New York State Alternate Assessment Technical Report 2014–15

Submitted to:

The New York State Education Department

Office of State Assessment

Prepared by:

measured progress

100 EDUCATION WAY, DOVER, NH 03820 (800) 431-8901
WWW.MEASUREDPROGRESS.ORG

TABLE OF CONTENTS

CHAPTER 1 OVERVIEW

This technical report provides an overview of the New York State Alternate Assessment (NYSAA), including a description of the purpose of the NYSAA, the processes utilized to develop and implement the NYSAA program, and Stakeholder involvement in those processes. By comparing the intent of the NYSAA with its process and design, the validity of the assessment's use can be evaluated. Starting with the 2013–14 NYSAA, a new test design was implemented for all content areas. For the 2013–14 NYSAA, new English Language Arts (ELA) and mathematics Extensions, which are aligned to the Common Core Learning Standards (CCLS), were developed. For science and social studies, Alternate Grade-Level Indicators (AGLIs) aligned to the New York State Learning Standards, development occurred in 2006–07 and 2007–08. The processes for developing the Extensions and AGLIs are presented in detail. Stakeholder input in the development of the overall NYSAA process itself is described in detail, including the content alignment of the Extension and AGLI design following the new Blueprint/test design, the Assessment Task development, the teacher trainings for administration, the Scoring Trainings and process, and the standard setting.

1.1 PURPOSE OF THIS REPORT

The purpose of this report is to document the technical aspects of the 2014–15 NYSAA. During the 2014–15 school year, approximately 22,226 students in Grades 3 through 8 and in high school were administered the NYSAA. ELA and mathematics were assessed at the Grades 3 through 8 and high school levels; science was assessed at the Grades 4 and 8 and high school levels; and social studies was assessed at the high school level.

Several technical aspects of the NYSAA are described in an effort to contribute to evidence supporting the validity of NYSAA score interpretations. Because the interpretations of the test scores are evaluated for validity, not the test itself, this report presents documentation to substantiate intended interpretations (AERA et al., 2014). Each chapter in this section contributes important information regarding the test's validity by addressing one or more of the following aspects of the NYSAA: Extension and AGLI and Assessment Task development, alignment, administration, scoring, reliability, standard setting, and achievement levels.

Standards for Educational and Psychological Testing (AERA et al., 2014) provides a framework for describing sources of evidence that should be considered when constructing an argument for assessment validity. These sources are found in five general areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. Although each of these sources may speak to a different aspect of validity, they are not distinct types of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

1.2 Organization of This Report

This report is organized based on the conceptual flow of the NYSAA as a multi-year-long process, which includes Blueprint design/development (completed in 2012–13), Extension (completed in 2012–13) and AGLI (completed in 2006–07 and 2007–08) development, Assessment Task development (completed in 2012–13), administration (completed in 2014–15), scoring (completed in 2014–15), standard setting (completed in 2013–14), technical characteristics, and validity. The appendices contain supporting documentation.

1.3 Current Year Updates

The NYSAA was redesigned, beginning with the 2013–14 administration. There were no changes from the 2013–14 administration to the 2014–15 administration. The overall structure of the NYSAA remains consistent with past practice in that it is a datafolio-style assessment that includes student performance data and evidence, and it is designed to assess students with severe cognitive disabilities.

CHAPTER 2 THE STATE ASSESSMENT SYSTEM

In New York State, both the general large-scale assessments and the alternate assessment test students on English Language Arts (ELA) and mathematics curriculum content taught during Grades 3 through 8 and high school; on science content taught during Grades 4 and 8 and high school; and on social studies content taught during high school. All students participate in the statewide assessment program through the following: the general assessments, with or without accommodations; the alternate assessment, with or without accommodations; or a combination of the general and alternate assessments, with or without accommodations.

2.1 INTRODUCTION

The New York State Alternate Assessment (NYSAA) is designed to provide a snapshot of an individual student's performance. A broader picture will emerge as the student results on the NYSAA are reviewed, along with results on other classroom and district assessments.

The NYSAA is a datafolio-style assessment that measures how well students with severe cognitive disabilities meet the standards at alternate achievement levels. All students, including those with severe cognitive disabilities, are required by federal law to have access to the general education curriculum. The New York State Education Department (the Department) has aligned Extensions with Common Core Learning Standards (CCLS) in ELA and mathematics, and Alternate Grade-Level Indicators (AGLIs) with the core curriculum in science and social studies, for the administration of the NYSAA. The content-area subject matter assessed by the NYSAA is clearly related to the grade-level content. While the content is reduced in scope and complexity, students with severe cognitive disabilities are held to high expectations in order to achieve the standards. The Extensions and AGLIs afford students a richer learning experience.

School districts across the United States are required to assess all students according to federal statute and state regulations. Assessment results tell educators how students are progressing and signal where changes may need to be made in the curriculum and/or instruction at the district, school, and classroom levels. Teachers should assess students in all areas (academic, social, etc.) on an ongoing basis, as part of the instruction cycle.

The No Child Left Behind (NCLB) Act of 2001 and the NYSAA are, in part, designed to raise expectations for students' academic achievement. Students with severe cognitive disabilities, when given the appropriate instruction and access to the general education curriculum, have demonstrated progress in their knowledge, skills, and understanding in academic content areas that were not initially anticipated by school personnel or parents. Higher expectations require that students with severe

cognitive disabilities have access to the general education curriculum and be provided with specialized instruction, as well as participate in national, state, and local assessment programs.

The administration period for the 2014–15 NYSAA was September 29, 2014 through February 27, 2015. The scoring period for the 2014–15 NYSAA was March 16, to May 1, 2015. The general sequence of events for administering the NYSAA is summarized below.

Summary of NYSAA Events

1. Each student's Committee on Special Education (CSE) determines how a student participates in the New York State Testing Program. The CSE uses the Department's guidelines regarding eligibility and participation criteria to guide their decision-making.
2. For each content area assessed, the student's instructional team, headed by the Lead Special Education Teacher (teacher), considers the most appropriate level of complexity for the student in each content area assessed. Five Extensions are required in ELA and mathematics, and two AGLIs are required in science and social studies.
3. Parents/guardians meet with the teacher to discuss how the NYSAA is administered and which specific Extensions and AGLIs will be used to assess their child.
4. Members of the student's instructional team conduct the baseline data point early in the administration period and document and rate student performance. Based on the results of the baseline assessment, the teacher will determine whether it is necessary to select another. The baseline data point serves two purposes: first, to confirm that the appropriate Level of Complexity has been selected and, second, to confirm that the student has not already mastered the selected skill. The baseline score cannot be higher than 74%.
5. Once the baseline administration confirms the task to be assessed, the instructional team provides instruction on the assessed skill, continuing to evaluate student progress until it appears that the student has acquired the skill.
6. Following the instructional period, a final data point is administered and scored for Level of Accuracy. The date of the final data point should not be less than 15 school days after the date of the baseline data point, and should occur as close to the end of the administration period as possible (no later than February 27, 2015). Similar items and materials should be used for both the baseline and final administrations.
7. The teacher assembles a datafolio containing the evidence of student performance and the percentages of the student's Level of Accuracy. The completed datafolio is submitted to the building administrator on or before the last day of the administration period, who then ships it to the regional Scoring Institute.
8. The NYSAA datafolios are scored at regional NYSAA Scoring Institutes during the scoring period defined by the Department.
9. Student reports are created and are made available to school districts, teachers, and parents/guardians.

2.2 Alternate Assessment Based on Alternate Achievement Standards

Up to 1% of New York State students in the grades tested may show academic proficiency through administration of an alternate assessment that is based on alternate achievement standards. The NYSAA is designed for those students with such severe cognitive disabilities that they are unable, even with the best instruction and appropriate accommodations, to participate in a general New York State assessment. The NYSAA is designed under the guiding philosophy that alternate achievement standards are built upon measurable, targeted skills linked to the CCLS in ELA and mathematics and to the core curriculum's Grade-Level Indicators in science and social studies. However, the alternate achievement standards represent student performance at lower levels of breadth, depth, and complexity than those found in the general assessments.

2.3 The Alternate Assessment System

The Individuals with Disabilities Education Act of 1997 (IDEA of 1997) requires that students with disabilities be included in each state's system of accountability and have access to the general curriculum. The federal reauthorization of the Elementary and Secondary Education Act, known as the NCLB Act of 2001, also speaks to the inclusion of all children in a state's accountability system by requiring states to report achievement for all students, as well as for groups of students on a disaggregated basis. These federal laws reflect an ongoing concern about equity: All students need to be academically challenged and taught to high standards. It is also necessary that all students be involved in the educational accountability system. Alternate achievement standards are reduced in breadth, depth, and complexity, but are linked to the same general curriculum standards taught to all students.

The IDEA of 1997 and the NCLB Act of 2001 clearly outline that all students, regardless of disability, participate in a statewide assessment system and be held accountable to the state standards. The NYSAA was developed to meet the requirements of these federal mandates; to provide a technically sound method to observe and record student achievement; to represent the breadth and depth of statewide content; to promote access to the general curriculum; to provide critical information to the CSE for use in the development of Individualized Education Programs (IEPs); and to meet criteria for alignment, access, burden, bias, sensitivity, and age appropriateness for students with severe cognitive disabilities. The 2013–14 NYSAA was the first year of administration of the new test design linked to the Common Core Learning Standards for ELA and mathematics. This same test design was used for the 2014–15 NYSAA administration.

2.4 Purpose of the Alternate Assessment System

The NYSAA measures the achievements of students with severe cognitive disabilities relative to the New York State Learning Standards, using alternate achievement levels based on a datafolio approach (as described in the next section). To ensure that this student population has access to the general education curriculum, for the NYSAA administration, the Department aligned the Extensions and AGLIs (discussed in the next section) with the CCLS in ELA and mathematics and with the core curriculum's Grade-Level Indicators in science and social studies.

The NYSAA is, in part, designed to raise expectations for students' academic achievement. Experience has shown that students with severe cognitive disabilities, when given appropriate instruction and access to the general education curriculum, demonstrate unanticipated progress in their knowledge, skills, and understanding in academic content areas. Prior to 2006–07, access to the general education curriculum was not necessarily a part of instructional programs for students with severe cognitive disabilities. In a recent survey of teachers who administered the NYSAA in 2014–15, 56% agreed that the Extensions and AGLIs assessed in the NYSAA made the grade-level core curricula more accessible, and said that the AGLIs are used in planning daily instruction.

The process for assessing the academic achievements of students who have severe cognitive disabilities and who are eligible for the NYSAA is outlined through structured guidelines and steps in the *2014–15 NYSAA Administration Manual* (accessible at http://www.p12.nysed.gov/assessment/nysaa/nysaa-manual-15.html). The process of datafolio development (see Chapter 7) supports the procedural validity of assessing students with severe cognitive disabilities, while being flexible enough to meet each individual student's learning needs and modalities.

2.5 Test Use and Decisions Based on Alternate Assessment

New York State conducts a statewide assessment program on an annual basis for all students in Grades 3 through 8 and high school. The NYSAA ensures that students with severe cognitive disabilities are included in the New York State Testing Program and that their results are included in all Adequate Yearly Progress (AYP) determinations.

The NYSAA is a datafolio-style assessment based on the assessment of Extensions and AGLIs. A datafolio is a collection of evidence of a student's academic performance, which is compiled by the student's instructional team and scored by qualified Scorers. By gathering performance data, the instructional team can provide parents/families/guardians and the CSE with an understanding of the student's knowledge, skills, and understanding as they relate to the CCLS in ELA and mathematics, and the New York State core curriculum in science and social studies. The CSE can use the datafolio to understand the student's achievement relative to these standards and to contribute to the

development of the student's IEP. Datafolios are scored during a standardized scoring period each spring. The NYSAA student reports are generally available in the fall following administration.

Performance levels, based on alternate academic achievement standards, were developed through a rigorous standard-setting process in June 2014. Alternate Performance Level Descriptors (APLDs) that outline the knowledge, skills, and understanding that a student may demonstrate within each grade and content area were edited and refined by panelists during the standard-setting process. The APLDs and datafolios, provide information to parents/families/guardians, the CSE, and the instructional team regarding potential modifications or adjustments to the student's instructional program.

2.6 Background and General Format

A datafolio is a collection of evidence of a student's academic performance compiled by the student's instructional team and scored by qualified Scorers. Instructional team members document student performance by recording the student's Level of Accuracy percentage as he or she performs an Assessment Task on two different dates, a baseline data point and a final data point, within the administration period. To verify this documentation, each datafolio must include student work products, Data Collection Sheets, photographs, or digital video and/or audio recordings. Teachers complete the required forms and submit all documentation and evidence in a binder or fastened folder for regional scoring.

Teachers are provided with a *NYSAA Administration Manual* that outlines the assessment requirements and the steps for compiling a datafolio, and includes the documentation forms and the NYSAA Frameworks as appendices. The NYSAA Frameworks include an introduction, and the NYSAA Test Blueprints outline the standards that will be assessed via the alternate assessment for each grade. The Test Blueprints illustrate, for each content area (e.g., ELA, mathematics, science, and social studies), the major areas of the standards' focus that teachers must assess at each grade. For both ELA and mathematics, five standards are assessed for each student; in science and social studies, two standards are assessed. In ELA and mathematics, teachers select an Extension from one of three Levels of Complexity, based on their students' needs. For students taking the NYSAA in science (Grades 4 and 8 and high school) and social studies (high school), teachers select an AGLI from one of three Levels of Complexity.

Teachers must identify one Extension or AGLI, based on the student's assessed grade level. The grade level corresponds to the student's date of birth. For each Extension or AGLI, the teacher must collect and document student performance data from an Assessment Task administered on two separate dates—the baseline and final. Once the baseline is administered, the teacher continues to provide instruction and evaluate student progress until he or she reaches a point where performance plateaus or the end of the administration period occurs. At this point, the final administration is

conducted. The administration guidelines recommend that there be at least 15 school days between the baseline and final administrations, which allows for instruction and for the student to learn the new skill. More than 15 days is acceptable, as long as the final administration takes place prior to the end of the administration period. One piece of verifying evidence must be submitted to demonstrate the student performance for each of the documented data points (baseline and final).

2.7 TESTING ACCOMMODATIONS

The CSE determines whether a student will participate in the alternate assessment with or without accommodations. Guidelines regarding accommodations are provided in the *NYSAA Administration Manual.* The CSE determines which testing accommodations are required, based on the student's documented needs. Testing accommodations:

- are consistent with the student's IEP;
- are designed to allow the student to demonstrate his or her knowledge, skills, and understanding with greater independence;
- do not change the level of the assessment, the construct of the assessment, or the criteria of the Assessment Task; and
- are provided to the student during instruction and not just for assessment.

For more information on testing accommodations, refer to Test Access and Accommodations for Students with Disabilities: Policy and Tools to Guide Decision-Making and Implementation (May 2006) at www.p12.nysed.gov/specialed/publications/policy/testaccess/policyguide.htm.

Frequently asked questions about testing accommodations and the NYSAA can be found at www.p12.nysed.gov/assessment/nysaa/home.html.

CHAPTER 3 THE STUDENTS

New York State conducts a statewide testing program on an annual basis for all students in Grades 3 through 8 and high school. The New York State Alternate Assessment (NYSAA) is a part of this statewide testing program. Designed for students with severe cognitive disabilities, the NYSAA measures student progress toward meeting the learning standards established for all students in the academic content areas of English Language Arts (ELA), mathematics, science, and social studies. The NYSAA ensures that students with severe cognitive disabilities are included in the State Testing Program (NYSTP) and that their results are accounted for as required by the No Child Left Behind (NCLB) Act of 2001 and the Individuals with Disabilities Education Act (IDEA) of 1997.

3.1 TARGET POPULATION

The target population for the NYSAA is extremely specific, and participation is limited to students with severe cognitive disabilities. The eligibility and participation criteria provide a definition of a student with a severe disability in accordance with Section 100.1 of the Regulations of the Commissioner of Education. For reference, this information is provided in the *NYSAA Administration Manual* and on the Web site of the New York State Education Department (the Department).

"Students with severe disabilities" refers to students who have limited cognitive abilities, combined with behavioral and/or physical limitations, and who require highly specialized educational and/or social, psychological, and medical services in order to maximize their full potential for useful and meaningful participation in society, and for self-fulfillment. Students with severe disabilities may experience severe speech, language, and/or perceptual-cognitive impairments and challenging behaviors that interfere with learning and socialization opportunities. These students may also have extremely fragile physiological conditions and may require personal care, physical/verbal supports, and assistive technology devices.

The process of determining eligibility begins with the Committee on Special Education (CSE). The CSE determines, on an individual basis, whether the student will participate in:

- the State's general assessment, with or without accommodations;
- the State's alternate assessment, with or without accommodations; or
- a combination of the State's general assessment for some content areas and the State's alternate assessment for other content areas.

The CSE ensures that decisions regarding participation in the State Testing Program are not based on:

- category of disability,
- language differences,
- excessive or extended absences, or
- cultural or environmental factors.

The CSE also ensures that each student has a personalized system of communication that addresses his or her needs regarding disability, culture, and native language so that the student can demonstrate his or her present level of performance. Tests and other assessment procedures are conducted according to the requirements of Section 200.4(b)(6) of the Regulations of the Commissioner of Education and Section 300.320(a)(6) of the Code of Federal Regulations.

Only students with severe cognitive disabilities are eligible for the NYSAA. The CSE determines whether a student with a severe cognitive disability is eligible to take the NYSAA, based on the following criteria:

- the student has a severe cognitive disability and significant deficits in communication/language and significant deficits in adaptive behavior; and
- the student requires a highly specialized educational program that facilitates the acquisition, application, and transfer of skills across natural environments (home, school, community, and/or workplace); and
- the student requires educational support systems, such as assistive technology, personal care services, health/medical services, or behavioral intervention.

While the State Testing Program provides full access to all students, 1% of students with severe cognitive disabilities in Grades 3 through 8 and high school are alternately assessed and are counted as proficient for purposes of accountability.

In accordance with 34 CFR 200.13 Adequate Yearly Progress in General, there is a 1% cap on the number of proficient and advanced scores on the alternate assessment that may be included in Adequate Yearly Progress (AYP) calculations at both the State and district levels.

3.2 Summary of Participation Rates

Tables 3-1 through 3-4 show a summary of participation in the 2014–15 NYSAA by demographic category for each content area.

Table 3-1. 2014–15 NYSAA: Summary of Participation—English Language Arts

Demographic Group	*Number Tested*	*Percent Participation*
All Students	21,710	100.00
Male	14,773	68.05
Female	6,937	31.95
American Indian/Alaskan Native	190	0.88
Black	5,407	24.91
Asian	1,176	5.42
Hispanic	5,842	26.91
White	8,779	40.44
Native Hawaiian/Other Pacific Islander	85	0.39
Multi	231	1.06

Table 3-2. 2014–15 NYSAA: Summary of Participation—Mathematics

Demographic Group	*Number Tested*	*Percent Participation*
All Students	21,716	100.00
Male	14,771	68.02
Female	6,945	31.98
American Indian/Alaskan Native	190	0.87
Black	5,409	24.91
Asian	1,177	5.42
Hispanic	5,841	26.90
White	8,784	40.45
Native Hawaiian/Other Pacific Islander	84	0.39
Multi	231	1.06

Table 3-3. 2014–15 NYSAA: Summary of Participation—Science

Demographic Group	*Number Tested*	*Percent Participation*
All Students	9,231	100.00
Male	6,183	66.98
Female	3,048	33.02
American Indian/Alaskan Native	74	0.80
Black	2,324	25.18
Asian	492	5.33
Hispanic	2,364	25.61
White	3,847	41.67
Native Hawaiian/Other Pacific Islander	39	0.42
Multi	91	0.99

Table 3-4. 2014–15 NYSAA: Summary of Participation—Social Studies

Demographic Group	*Number Tested*	*Percent Participation*
All Students	2,886	100.00
Male	1,875	64.97
Female	1,011	35.03
American Indian/Alaskan Native	26	0.90
Black	731	25.33
Asian	141	4.89
Hispanic	657	22.77
White	1,298	44.98
Native Hawaiian/Other Pacific Islander	10	0.35
Multi	23	0.80

CHAPTER 4 TEST DEVELOPMENT

4.1 FRAMEWORKS OF THE TESTING PROGRAM

The New York State Common Core Learning Standards (CCLS) provide the framework for the New York State Testing Program in English Language Arts (ELA) and mathematics. The State's core curriculum learning standards provide the framework for the New York State Testing Program in science and social studies. Each statewide assessment program has a Test Blueprint that outlines the priorities to be assessed based on the grade-level learning standards. The redesign of the New York State Alternate Assessment (NYSAA), which began with the 2013–14 administration, was done in response to changes that the New York State Education Department (the Department) made to the general education assessments to assess the CCLS in ELA and mathematics. The general education assessment Blueprints were used as the basis for the development of the alternate assessment Test Blueprints, which, in turn, drove the alternate assessment content. There is one alternate assessment Blueprint for each of the four content areas assessed (see Appendix A).

In May 2012, the Department assembled teacher committees to review the Test Blueprints for ELA and mathematics. The group's goal was to develop Essences and Extensions for each standard. Groups focused on designing Extensions that aligned to general education grade-level content and, most importantly, were appropriate for students with severe cognitive disabilities. The draft Essences were reviewed by the Department and Measured Progress, and then posted on September 10, 2012 for public comment through October 5, 2012. A total of 852 respondents began the survey and 66.7% completed the survey. To the greatest extent possible, feedback collected from the survey was incorporated into the Essence and Extension documents.

In October 2012, the groups were reassembled with the purpose of developing Assessment Tasks for the approved Extensions. The draft Assessment Task documents were posted on December 7, 2012 for public comment through January 4, 2013. A total of 1,026 respondents began the survey and 60.3% completed the survey. To the greatest extent possible, feedback collected from the survey was incorporated into the Assessment Task documents.

The Department followed a similar process in fall 2006, when it assembled special education and general education teacher committees to review the core curricula and general education assessment Blueprints for science and social studies. This group's goal was to determine academic content priorities for the NYSAA, based on the core curricula, general education assessment Blueprints, and, most importantly, applicability for students with severe cognitive disabilities. The

process was designed to ensure alignment with general education grade-level content and to promote higher expectations for students taking the NYSAA.

The special education and general education teacher committees' discussions focused on the actual depth and breadth of the alternate assessment requirements. Throughout the review, psychometricians from the Department and Measured Progress provided direction for maintaining a valid and reliable assessment. The resulting work by the special education and general education teacher committees expanded the standards for students with severe cognitive disabilities and created Extensions for ELA and mathematics, and Alternate Grade-Level Indicators (AGLIs) for science and social studies. The Extensions and AGLIs provide entry points to the grade-level content of the standards so that a student's level can be gauged in terms of the standards established for all students by the New York State Board of Regents.

The Test Blueprints, CCLS and core curriculum standards, Essences, Extensions and AGLIs, and Assessment Tasks for each grade can be found in the *2014–15 NYSAA Administration Manual*: Appendix F—NYSAA Frameworks (http://www.p12.nysed.gov/assessment/nysaa/nysaa-manual-15.html).

4.2 Extensions and AGLIs Mapped to NYS Learning Standards and Core Curriculum by Grade

The Extensions are aligned to the State's CCLS, and AGLIs are aligned to the New York State learning standards. Both the Extensions and AGLIs reflect high expectations for students with severe cognitive disabilities. This alignment is illustrated in Figure 4-1.

For the Extensions, a teacher committee meeting was held in October 2012 to review the new Test Blueprints for ELA and mathematics, and to develop Essences and Extensions. A second meeting was held in May 2013 for the purpose of expanding the Extensions into Assessment Tasks.

For the AGLIs, teacher committee meetings were held during the summer and early fall of 2006 to gather input on aligning the NYSAA requirements with Grade-Level Expectations and on developing AGLIs. Additionally, teacher committee meetings were held in spring 2007 and 2008 to further refine the AGLIs and to develop additional Assessment Tasks for teachers to use in the alternate assessment. As part of the overall redesign implemented in 2013–14, the science and social studies AGLIs were narrowed and the Assessment Tasks were updated to follow the same format and philosophical approaches as the Extensions and Assessment Tasks in ELA and mathematics.

The Board of Regents approved a set of learning standards to guide instruction and assessment. The learning standards serve as the basis of the core curricula in science and social studies. The curriculum of each content area is divided into the following components:

- science: standards and key ideas

- social studies: standards and units

Each component in a content area lists Grade-Level Expectations for student performance. These expectations are called grade-level performance indicators, or content understandings.

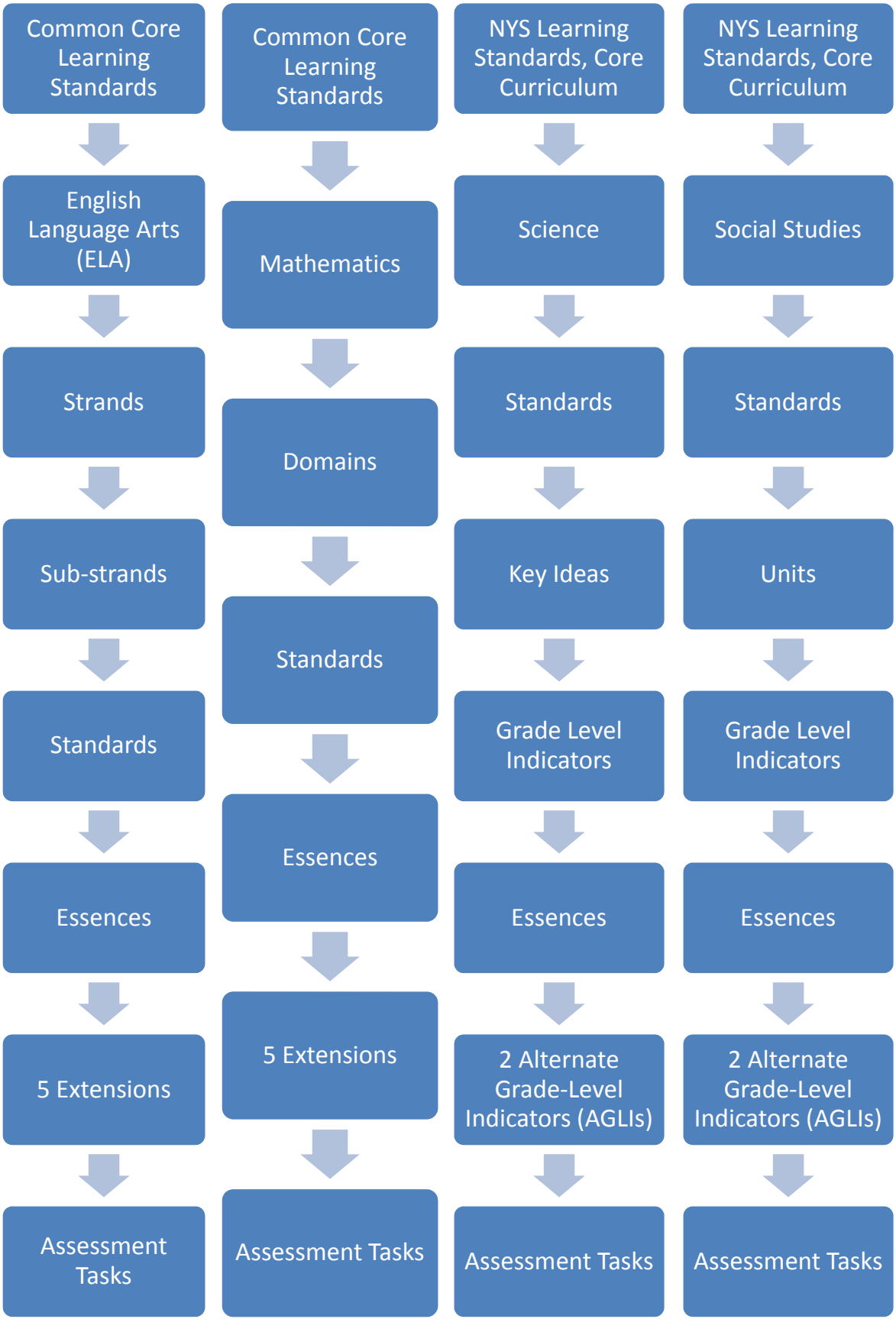
Grade-Level Expectations are further distilled into essences. Essences are the “big ideas” of the Grade-Level Expectations for a grade. Assessment is based on the Essences for each component of each content area. AGLIs are aligned to the Essences in terms of three Levels of Complexity.

The Board of Regents approved the New York State P-12 CCLS for ELA and mathematics. The New York State Learning Standards serve as the basis of the core curricula in ELA and mathematics. The curriculum of each content area is divided into the following components:

- ELA: strand and sub-strands
- mathematics: domain

Each component in a content area lists Grade-Specific Standards, which are further distilled into Essences. Assessment is based on the Essences for each component of each content area. Extensions are aligned to the Essences in terms of three Levels of Complexity.

Figure 4-1. 2014–15 NYSAA: Mapping of Assessment Tasks to the Learning Standards



4.3 AGLI Selection Criteria and Process

The New York State Board of Regents committed to the Common Core State Standards (CCSS) in January 2010 and formally adopted the CCSS for ELA and mathematics in July 2010. The New York State P-12 Common Core Learning Standards (CCLS) incorporated State-specific additions in January 2011. The Board of Regents announced that, for students with severe cognitive disabilities, student progress on the CCLS would be measured beginning with the 2013–14 administration of the NYSAA in ELA and mathematics. Beginning in November 2011, the Department and Measured Progress began developing a new test design aligned to the CCLS. In this design, new Extensions would be developed for ELA and mathematics, and the existing AGLIs would be refined in science and social studies. The process for developing the Extensions and previously the AGLIs is outlined below.

Extensions:

The Department coordinated the recruitment of teacher committees, which met in May and October 2012. Participants were chosen by the Department with the intent that the participants would remain consistent across both meetings, which ensured consistency in the overall process and content interpretation.

Participants were assigned to grade-content workgroups. Each group reviewed the CCLS and the new Test Blueprints with the purpose of developing Essence statement(s) for each standard being assessed. Once the Essences were developed, the panels worked to create three Extensions for each standard, one for each Level of Complexity (less, middle, and more complex). The following expected outcomes were provided to the work groups:

1. Each grade-content workgroup will produce a final draft version of an Essence statement that addresses the emphasis of the standard (ELA) or cluster (mathematics) and an Extension(s) at three complexity levels (less, middle, and more complex) following the draft NYSAA Test Blueprint. The Essence and Extensions will be determined by considering:
    - Curricular congruence and alignment;
    - Developmental applicability for students with significant cognitive disabilities and datafolio product alignment and feasibility;
    - Applicability to transitional and career readiness skills for students with significant cognitive disabilities; and
    - Parental and special populations experiences ensuring consideration of all variations of abilities of students with significant cognitive disabilities.
2. Workgroups will have in-depth discussions between special education and general education teacher committees on the standards (ELA) or clusters (mathematics), Essence

statements, and Extension statements to develop the final draft version of the Essences and Extensions.

During an opening session facilitated by the Department and Measured Progress, participants were welcomed and introduced. An overview of the process and the format of the materials were presented. Following the opening session, participants moved into their assigned grade-content work groups. Using a standardized template, each group was asked to follow the same basic steps for their work.

Step 1: Introductions and Material Review.

The participants in each grade-content workgroup introduced themselves and indicated which region they were representing. A room facilitator reviewed the expectations for their work and identified a note taker to record key points of their discussions and decisions. Participants were asked to familiarize themselves with the layout of the CCLS documents and the NYSAA Alignment to CCLS template.

Step 2: Develop Essence statement(s).

Using the Alignment template as a guide, each workgroup considered the standard being assessed and developed one or more Essence statements. These statements narrowed the depth and breadth of the content, which enables students with significant cognitive disabilities to access the content and demonstrate their knowledge, skills, and understanding.

Step 3: Develop Extensions aligned to the Essence statement and transition skills.

Using the Essence statement(s), the workgroups developed three Extensions, which represented increasing complexity and cognitive demand. In addition, participants were asked to consider the Career Development and Occupational Studies standards and identify links to the CCLS.

Step 4: Review the group work.

Within each content area, workgroups shared the Essences and Extensions that they drafted regarding the progression of knowledge, skills, and understanding.

Following the workgroup meetings, an extensive review of the draft documents was conducted by content experts from the Department and Measured Progress. During the summer of 2012, the draft documents were posted on the Department's Web site for public comment. Based on the public comment, additional work was done on the Essences and Extensions before they were presented to the workgroups again in October 2012.

AGLIs:

The Stakeholder groups who met in 2006, 2007, and 2008 were named the NYSAA Revision Workgroup (NRWG). The participants who were chosen for the initial group remained throughout all of the NRWG meetings, which ensured consistency in the overall process and in content interpretation.

As part of the implementation of the new test design, the Test Blueprints for science and social studies were revised to narrow the content assessed. In addition, minor editorial revisions were made to the AGLIs and Assessment Tasks. However, as was the case with the 2012–13 version of the NYSAA Frameworks, the intent of the AGLIs was not changed in any way.

The spring 2008 NRWG process was consistent in science and social studies. The NRWG was not allowed to edit or change the Test Blueprints, Grade-Level Expectations, Essences, and intent of the AGLIs. As outlined below, for each content area, three steps were followed by the participants, and the fourth step was completed afterward by the content developers.

Step 1: Present the expected outcomes for the workgroup.

The workgroup was welcomed and thanked for participating in the revision of the NYSAA Frameworks. The participants introduced themselves and indicated where they were from and in which content area they were participating. The presentation then consisted of directing the groups through the materials that they would be working with and explaining the specific tasks for the grade-content workgroups, as well as other logistical information. The workgroup was given time for questions and then released into their grade-content workgroups, where they remained for the rest of the day and the following day.

Step 2: Review the Frameworks and other materials.

In order to complete the tasks required in the time allotted, each content area facilitator divided participants into groups by grade level and distributed the materials for review. The groups were divided as indicated in Table 4-1.

Table 4-1. 2014–15 NYSAA: NRWG Participant Groups from 2008

Content Area	*Group*	*Grades*
Science	1	4
	2	8
	3	High School
Social Studies	1	5
	2	8, High School

Step 3: Complete the work process.

In all the content area workgroups, the participants reviewed and edited existing Sample Assessment Tasks (SATs) and then worked to add new SATs. The process for adding new SATs was as follows: The workgroups first focused on AGLIs that did not have SATs. Then they developed additional SATs for AGLIs that already had at least one SAT. Throughout the editing and developing of SATs, each workgroup worked to ensure alignment to the AGLIs. During the editing process, the workgroups also identified words that they felt should be added to the glossary for each content area. The tasks within each content area focused on each of the outcomes identified from the revision of the NYSAA Frameworks.

Step 4: Review the group work as a further check on core curriculum alignment.

Each facilitator gathered each workgroup's work and reviewed all edits and suggestions, as another check on content alignment. The edited NYSAA Frameworks then went to the Department for an additional content-alignment check and for finalization of each content area for the 2008–09 administration of the NYSAA.

4.4 ASSESSMENT TASK DEVELOPMENT

In October 2012 the Essence/Extension workgroups were reassembled with the purpose of developing Assessment Tasks aligned to the Extensions. Their process was similar to the steps followed during the May 2012 meeting. Participants were not allowed to edit or revise the Essences or Extensions. Using an updated Alignment template, the groups began with the first standard in their assigned grade-content and developed at least one Assessment Task aligned to each Extension. For both the Extensions and AGLIs, an Assessment Task describes an observable student action related to the specific knowledge, skills, and understanding aligned to the AGLI and, in turn, to the core curriculum.

The 2008 NRWG developed, edited, and refined the original Assessment Tasks aligned to the AGLIs. Regional Lead Trainers (RLTs), who were part of the NRWG, provided input on SATs aligned to the AGLIs. Teachers had the opportunity to submit assessment tasks for possible inclusion in the NYSAA Frameworks through the annual online teacher survey. Information collected during the 2011–12 administration and scoring periods also influenced edits to the SATs. Edited SATs were reviewed and approved by the Department for the 2012–13 NYSAA Frameworks. See the following section for more information on task development and refer to the *NYSAA Administration Manual* for information provided to teachers regarding Assessment Task requirements.

CHAPTER 5 TEST CONTENT

The New York State Alternate Assessment (NYSAA) is intended to provide students with severe cognitive disabilities the opportunity to participate in a statewide assessment that is both meaningful and academically challenging. Given the wide diversity of this student population, great emphasis is placed on ensuring that Grade-Level Expectations within the Common Core Learning Standards (CCLS) for English Language Arts (ELA) and mathematics and the New York State learning standards for science and social studies are accessible to all students. The assessment design allows students to demonstrate their knowledge, skills, and understanding of the CCLS through Extensions for ELA and mathematics and the New York State Learning Standards through the Alternate Grade-Level Indicators (AGLIs) for science and social studies. The Extensions and AGLIs are organized into three Levels of Complexity to provide an appropriate entry point for students into the standards and to maintain the academic focus of the alternate assessment. Student performance data—Level of Accuracy—is collected by the teacher for each Extension and AGLI that the student is assessed against.

5.1 ALTERNATE PERFORMANCE LEVEL DESCRIPTORS (APLDS)

The APLDs, developed for the standard setting that took place in June 2014, were first used for the 2013–14 administration and reporting. The same APLDs were used for the 2014–15 administration and reporting. The purpose of the standard setting conducted in June 2014 was to establish cut scores for each alternate performance level in ELA and mathematics, Grades 3 through 8 and high school; in science, Grades 4 and 8 and high school; and in social studies, for high school.

The APLDs provided panelists with an idea of the knowledge, skills, and understanding related to the CCLS for ELA and mathematics and the core curriculum for science and social studies that a student at each of the four performance levels might demonstrate. A final activity during standard setting was for each group to provide suggestions for edits to the APLDs. The New York State Education Department (the Department) used the input to refine the APLDs for reporting. The APLDs are included in the NYSAA reports for districts, schools, parents/guardians, and educators to better explain each performance level.

5.2 ACCESS TO THE GENERAL CURRICULUM

The CCLS for ELA and mathematics contain grade-level content for pre-kindergarten through high school. Additionally, the core curricula for science and social studies contain grade-level content at the elementary, intermediate, and secondary levels. These core curricula are aligned with the New York State Learning Standards.

For the 2013–14 NYSAA, the Department, in cooperation with teacher committees from across the State, has expanded the CCLS in ELA and mathematics to Extensions for students with severe cognitive disabilities. Previously, the Department, in cooperation with special education teacher and general education teacher committees from across the state, had expanded the core curriculum Grade-Level Expectations in science and social studies to AGLIs for students with severe cognitive disabilities. Extensions and AGLIs provide an entry point to the grade-level content of the standards. Extensions and AGLIs measure a level of mastery of the knowledge, skills, and understanding aligned with the CCLS and core curricula established for all students by the New York State Board of Regents.

5.3 TEST FORMAT

The NYSAA is a collection of student work in the form of a datafolio. The NYSAA Test Blueprints outline for teachers the content to be assessed at each grade and content area combination. The NYSAA Test Blueprints for each content area are included in Appendix A. Each of the five content standards is required to be assessed for ELA and mathematics within each grade. Each of the two content standards is required to be assessed for science and social studies within each grade. ELA and mathematics are assessed in Grades 3 through 8 and high school. Science is assessed in Grades 4 and 8 and high school. Social studies is assessed in high school. Extensions and AGLIs are presented in the *NYSAA Administration Manual* in a spectrum of increasing complexity: less, middle, more. Extensions and AGLIs must be used as written. An Assessment Task is aligned to a specific Extension or AGLI. It describes the student action being assessed and outlines the basic expectation of what will be demonstrated in the verifying evidence. Teachers must use the Assessment Task as written, but, in most cases, there is more than one Assessment Task aligned to a specific Extension or AGLI that a teacher may select. Allowing teachers to select an Extension and AGLI and to then choose an Assessment Task aligned to that Extension or AGLI, results in individualization while maintaining the content consistency of the alternate assessment. Consistency is further ensured across grade levels and content areas by adherence to strict administration requirements for datafolios.

A datafolio is the resulting body of evidence of a student's academic performance across the content standards of selected Extensions or AGLIs, as compiled by the student's instructional team and scored by qualified Scorers. For each standard in ELA and mathematics, there are three Extensions presented across a spectrum of complexity from least to most complex. For each standard in science and social studies, AGLIs are presented across a spectrum of complexity from least to most complex. Teachers select the Extension or AGLI most appropriate for a student and conduct the assessment. Student performance is rated by the student's instructional team according to the student's Level of Accuracy in performing each Assessment Task. Two dates of student performance are documented for each standard in all content areas. Teachers first administer a baseline data point to collect performance data and evidence that confirm the student has not yet mastered the assessed skill.

Based on the outcome of the baseline data point, it may be necessary to adjust the Level of Complexity (choose another task at a higher or lower Level of Complexity). Following the baseline data point, teachers provide instruction and evaluate students to gauge student growth. Before the end of the administration period, a final administration is conducted and documented in the datafolio. In general, the Department recommends at least 15 school days between the baseline and final data points. To verify the baseline and final data points' documentation, each datafolio must include verifying evidence that demonstrates the student's performance of the task. Teachers may choose to submit the following as evidence: student work products, Data Collection Sheets, photographs, and/or digital video or audio recordings for the baseline administration performance and the final administration performance. Teachers complete the required forms, and submit all documentation and evidence in a binder or fastened folder for regional scoring. Detailed information about the content of and procedures for developing the datafolio is presented in the *NYSAA Administration Manual*.

5.4 ASSESSMENT DIMENSIONS

NYSAA datafolios are scored using two dimensions:

- Connection to Grade-Level Content

  The Connection to Grade-Level Content dimension is met when:

  - The Extension or AGLI is from the student's assessed grade;
  - the Assessment Task is clearly aligned with the Extension or AGLI; and
  - the verifying evidence submitted is aligned with the Assessment Task.

  Both of the connections must be clearly evident for the standard to be scored.

- Performance
  - Level of Accuracy is calculated as a percentage (0%–100%).

Level of Complexity is part of the NYSAA test design and, in addition to Level of Accuracy, factors into a student's overall performance level.

CHAPTER 6 ALIGNMENT

6.1 PROMOTING ALIGNMENT THROUGH ACHIEVEMENT LEVEL DESCRIPTORS

The Alternate Performance Level Descriptors (APLDs) for the New York State Alternate Assessment (NYSAA) are uniquely defined, by the use of unifying adverbs, for each grade and content area. The APLDs provide a structure for understanding the knowledge, skills, and understanding that a student may have demonstrated in their NYSAA datafolio at a performance level. The APLDs are meant to be a guide or a framework to give a picture of student performance. Due to the varying abilities of students with severe cognitive disabilities, the APLDs were developed to be a flexible definition of student performance on the NYSAA. The student performance documentation that is recorded and evidenced within the datafolio is a more prescribed and quantified system of documentation.

The development of APLDs occurred in 2014 as part of the NYSAA redesign. The APLDs for each grade and content area provided panelists participating in standard setting with the official description of the knowledge, skills, and understanding that students are expected to display for each performance level. The APLDs were developed by using the old APLDs and the general education Performance Level Descriptors for Grades 3 through 8 in English Language Arts (ELA) and mathematics. The initial language was developed by the Regional Lead Trainers (RLTs; see Chapter 7) and was then refined by Measured Progress. The APLDs were reviewed, edited, and approved by the Department.

The standard-setting panelists were able to come to a consensus with a generalized understanding of the terms described above due to their extensive knowledge of the NYSAA student population combined with understandings of the New York State Common Core Learning Standards (CCLS) for ELA and mathematics and the New York State core curricula for science and social studies. The APLDs provide information related to specific content assessed within a grade and content area and how that content skill may be performed by a student through his or her accuracy level. Many students who take the NYSAA have splinter skills, require various supports to perform, and can vary from day to day in their demonstration of the knowledge, skills, and understanding that they are working on within the datafolio. As such, the terms used within the APLDs provide some parameters and flexibility to allow for a basic picture of student performance without being specifically quantified. A set quantification would not allow for a representative understanding of a student with severe cognitive disabilities who took the NYSAA.

CHAPTER 7 ADMINISTRATION AND TRAINING

New York State utilizes a train-the-trainer model to provide training related to the New York State Alternate Assessment (NYSAA). Each Board of Cooperative Educational Services (BOCES) and Big Five City School District designates at least one person as an Alternate Assessment Training Network Specialist (AATN Specialist) and at least one person as a Score Site Coordinator (SSC). (The Big Five City School Districts are Buffalo, New York City, Rochester, Syracuse, and Yonkers.) AATN Specialists and SSCs participate in the regional Administration Training conducted in September and facilitated by the Department and Measured Progress. The AATN Specialist is responsible for conducting the NYSAA Administration Training with teachers. SSCs are responsible for the coordination of the regional Scoring Institutes; therefore, they also need to have an understanding of the NYSAA administration guidelines. In addition, nine Regional Lead Trainers (RLTs) provide technical assistance to assigned geographic regions across the state. The RLTs assist with administration and Scoring Training and Collegial Review processes, as well as provide support to teachers throughout the administration period via e-mail and telephone.

7.1 STEPS FOR ADMINISTRATION

The teacher, in coordination with the instructional team, is responsible for administering the NYSAA to a student. The *NYSAA Administration Manual* provides detailed guidelines on how to administer the NYSAA to a student. The NYSAA has a specific administration period during which the assessment can be conducted. Assessment data cannot be collected before or after the administration period. The administration period for 2014–15 was September 29, 2014, to February 27, 2015. The first step is to review the Individualized Education Program (IEP) for a student who has been designated to take the NYSAA and determine the grade that the student will be assessed at, using the student's date of birth and the NYSAA Age Range Chart. Next, the teacher determines the Extension or Alternate Grade-Level Indicator (AGLI) for each content standard on which the student will be assessed. For English Language Arts (ELA) and mathematics, five standards are assessed. For science and social studies, two standards are assessed. Then, the teacher determines an Assessment Task that will demonstrate the Extension or AGLI. The Assessment Task describes the student action being assessed. Once the Extensions or AGLIs and Assessment Tasks have been determined, the teacher conducts the Assessment Task with the student as a baseline administration. The baseline data point confirms that the student has not yet mastered the skill being assessed. If the student performance is 74% or below, then the teacher can continue to assess that skill. If the student performance is 75% or higher, then a higher-level skill must be assessed. If this is the case, the teacher would need to conduct a new baseline administration. Following the baseline administration, there is a period of instruction and

evaluation of the skill being assessed. Then, the teacher conducts the Assessment Task with the student as a final administration. The baseline data point and final data point administrations of the Extension or AGLI and Assessment Task are recorded and documented. Student performance includes the student's Level of Accuracy. Verifying evidence showing the student demonstrating the knowledge, skills, and understanding of the Extension or AGLI through the completion of the Assessment Task must be included for the baseline and final administration dates of student performance documented. There are four types of verifying evidence that can be included, each with specific guidelines on what must be included for it to be considered a valid piece of evidence at scoring. The four types are student work products, a sequence of captioned and dated photographs, digital video or audio clips, and Data Collection Sheets. Each datafolio is required to have at least one Collegial Review. Collegial Review requires a colleague or administrator who is familiar with the NYSAA, but is not the student's teacher who prepared the datafolio, to review the student's datafolio contents.

7.2 STEPS IN CONSTRUCTING THE DATAFOLIO

The *NYSAA Administration Manual* provides specific information on the construction and organization of the datafolio. For each Extension or AGLI, there must be a Data Summary Sheet. The Data Summary Sheet is the summarizing information regarding the Extension or AGLI. It includes student demographic information, the Extension or AGLI assessed, the Assessment Task, and student performance data. The baseline and final administration dates of performance data include the percentage for the Level of Accuracy. Also documented on the Data Summary Sheet for the baseline and final administrations is a *yes* or *no* to indicate if a student received verbal or physical cues or prompts to redirect or refocus the student on the Assessment Task. In addition to the individual requirements of each type of verifying evidence, the verifying evidence must include three required elements—student name, date of student performance, and Level of Accuracy. The teacher is responsible for ensuring that the verifying evidence connects to the Assessment Task and that it meets the requirements outlined in the *NYSAA Administration Manual,* in order to include it in the datafolio. On or before the end of the administration period, the teacher assembles the datafolio in a binder or fastened folder. The datafolio includes a NYSAA Student Page, which provides demographic information regarding the student and the grade assessed, supports required per the IEP, accommodations provided during testing, and the month that a Collegial Review was conducted. Although not required, a datafolio also includes a table of contents, which provides information to Scorers on where information is located in the datafolio. The ELA assessment documents come first, followed by mathematics, then science and social studies, if applicable. The Extensions or AGLIs within each content area are organized by using the numbers in the boxes in the upper-right corner of the Data Summary Sheets (Extension 1-Extension 5; AGLI 1-AGLI 2).

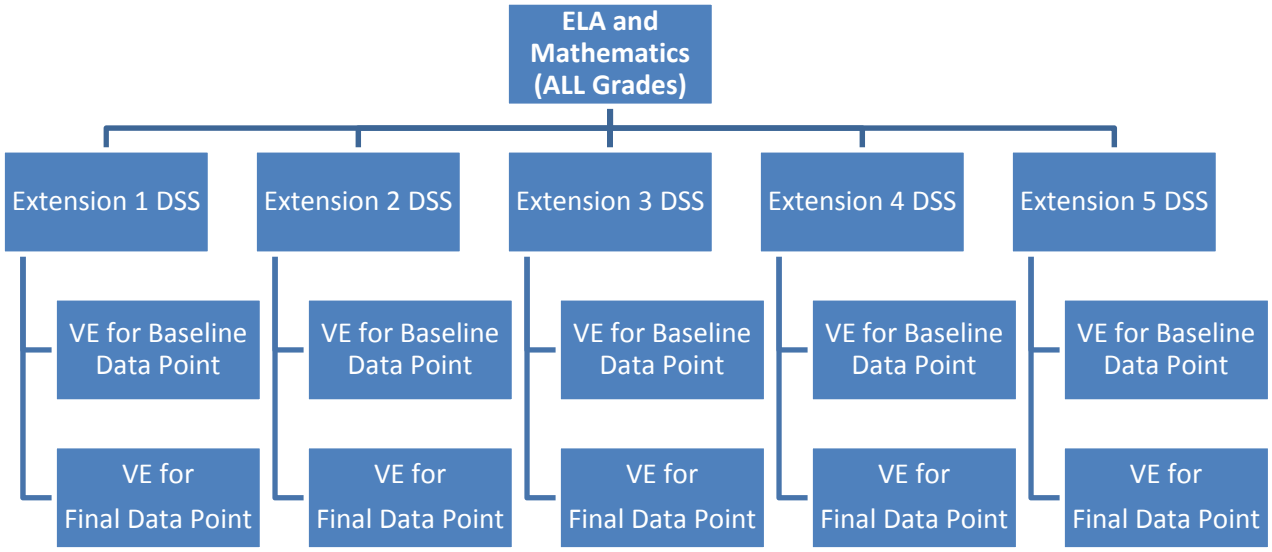
For ELA and mathematics, the order of the documents is as follows:

Extension 1:

- Data Summary Sheet
  - Verifying evidence for the baseline data point
  - Verifying evidence for the final data point
    - o If either piece of verifying evidence is a Data Collection Sheet, the supporting evidence directly follows the Data Collection Sheet.

This order is repeated for the remaining Extensions 2, 3, 4, and 5 in ELA and mathematics.

Figure 7-1. 2014–15 NYSAA: Datafolio Elements for ELA and Mathematics



For science and/or social studies, the order of documents is as follows:

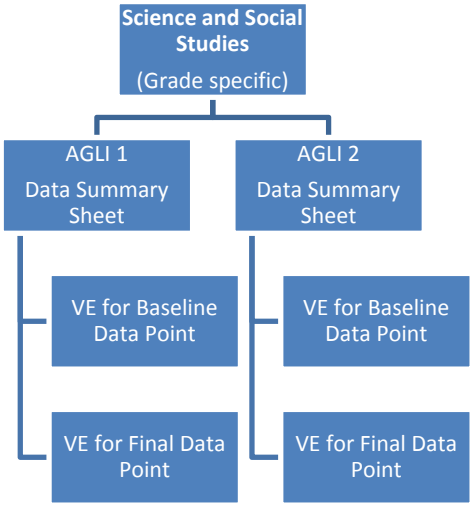
AGLI 1:

- Data Summary Sheet
- Verifying evidence for the baseline data point
- Verifying evidence for the final data point
  - o If either piece of verifying evidence is a Data Collection Sheet, the supporting evidence directly follows the Data Collection Sheet.

AGLI 2:

- Data Summary Sheet
- Verifying evidence for the baseline data point
- Verifying evidence for the final data point
  - o If either piece of verifying evidence is a Data Collection Sheet, the supporting evidence directly follows the Data Collection Sheet.

Figure 7-2. 2014–15 NYSAA: Datafolio Elements for Science and Social Studies



7.3 Administration Training and Collegial Review

In September 2014, the Department, in collaboration with Measured Progress, trained AATN Specialists and SSCs from across the state on how to conduct the NYSAA Administration Training with teachers in their regions. The one-day trainings were conducted regionally across the state over a two-week period. There were three main activities conducted. First, information regarding updates to the NYSAA and the materials was provided. Then, the NYSAA Administration Training DVD was shown. The training also included the completion and review of the Guided Practices. Last, the participants were asked to work in groups to discuss strategies to improve administration practices and how best to support teachers administering the 2014–15 NYSAA.

A total of five NYSAA Administration Trainings occurred at four geographically diverse sites: the Albany region, which included Long Island and the regions surrounding New York City; the Syracuse region; the Buffalo and Rochester region; and the New York City region, which included the non-District 75 trainers on one day and the District 75 trainers on another day. Table 7-1 outlines the number of participants at each training session.

Table 7-1. 2014–15 NYSAA: Administration Updates Training—Participant Count

	Albany Region	*Syracuse Region*	*Buffalo-Rochester Region*	*New York City Region (Two Trainings)*	*Total*
NYSAA Administration Updates Training	60	21	48	119	248

Administration Training for teachers is provided through a combination of Guided Practices and a training DVD. AATN Specialists are required to use all parts of the DVD and Guided Practices, as specified by the Department. The NYSAA Administration Training DVD is organized into segments.

There is an Opening segment; a Department Messages segment; a Steps for Administration segment; and a Best Practices, Recommendations, and Closing segment. An additional optional segment is provided, which is an introduction to an online tool that teachers can use to support their administration practice. The opening segment provides general information about what is going to be covered during the training session. The Department Messages segment provides the background for the implementation of the NYSAA test design and alignment of ELA and mathematics to the CCLS, and responses to several frequently asked questions related to the assessment process for the NYSAA. The Steps for Administration segment is a detailed review of each of the steps for administering the NYSAA, including things to consider in planning for the assessment, specifics regarding administering the assessment, and an outline of steps for assembling and submitting the datafolio for scoring. The information provided in this segment follows the organization of the *NYSAA Administration Manual,* and includes many visuals to assist teachers in understanding the NYSAA. The Best Practices, Recommendations, and Closing segment provides best practices tips and strategies on how to maintain the Connection to Grade-Level Content during administration, information on prompts and cues, and things to keep in mind for the 2014–15 NYSAA, as well as next steps for teachers and information regarding Collegial Reviews. At specific points throughout the segments, there are stop points built in, and a Guided Practice must be conducted. The Guided Practices reinforce the information that was contained in the segment. There are a total of four Guided Practices. The first Guided Practice focuses on understanding how to use a student's date of birth to determine the correct grade at which a student should be assessed, and how to use the Test Blueprints in the *NYSAA Administration Manual*: Appendix F—NYSAA Frameworks (http://www.p12.nysed.gov/assessment/nysaa/nysaa-manual-15.html). The second Guided Practice focuses on understanding how to navigate through the NYSAA Frameworks to select the Extensions and AGLIs and how to determine some verifying evidence options for specific Extensions and AGLIs. The third Guided Practice focuses on determining and documenting baseline student performance and determining if the baseline threshold is exceeded. The fourth Guided Practice provides teachers with a review of a NYSAA requirements review worksheet. Teachers complete all four Guided Practices, and a review of the practices is facilitated by the AATN Specialists. At or before the locally conducted NYSAA Administration Trainings, teachers are provided with the *NYSAA Administration Manual*, which includes the NYSAA Frameworks as Appendix F.

Collegial Review is required for each student datafolio. Collegial Review is an independent review of a datafolio. Reviewers should:

- be familiar with the current alternate assessment; and/or
- have attended the 2014–15 Administration Training in the fall of 2014.

The Department recommends that Collegial Reviews take place during the planning phase, at a midpoint during administration, and prior to the end of administration. The teacher is given feedback about whether the appropriate connections have been made between the Extensions or the AGLIs and the Assessment Tasks and between the Assessment Tasks and the verifying evidence. Also, Collegial Reviews help to confirm that all documents included in the datafolio at that point meet all procedural requirements. The Department cautions that a Collegial Review helps ensure, but does not guarantee, that a datafolio meets the procedural requirements necessary for a student to receive a reportable score.

CHAPTER 8 SCORING

Alternate Assessment Training Network Specialists (AATN Specialists) and Score Site Coordinators (SSCs) participate in the regional Scoring Training conducted each year. SSCs are responsible for the coordination of the regional Scoring Institutes, and must pass the qualification samples in order to make content decisions during Scoring Institutes. AATN Specialists act as Floor Managers at Scoring Institutes, and must also pass the qualification samples in order to make content decisions during Scoring Institutes.

In March 2015, the Department, in collaboration with Measured Progress, trained AATN Specialists and SSCs from across the State on how to score New York State Alternate Assessment (NYSAA) datafolios and how to conduct the NYSAA Scoring Training with Scorers at the Scoring Institute in their region. The one-day trainings were conducted regionally across the State over a two-week period. Three main activities were conducted. First, information regarding updates to the NYSAA Scoring Procedures, Decision Rules, and the scoring materials was provided. Then, the NYSAA Scoring Training DVD was shown. The training also included the completion and review of the Guided Practice samples. Last, participants were asked to complete a qualification process by scoring sample datafolios. Participants that did not meet the minimum performance requirements during the qualification process were retrained and provided with the opportunity to complete the qualification processes again with a new set of sample datafolios.

A total of five NYSAA Scoring Trainings occurred at four geographically diverse sites: the Albany region, which includes Long Island and the regions surrounding New York City; the Syracuse region; the Buffalo and Rochester region; and the New York City region, which included the non-District 75 trainers in one training session and the District 75 trainers in another training session. Table 8-1 outlines the number of participants at each training session.

Table 8-1. 2014–15 NYSAA: Scoring Training—Participant Count

	Albany Region	*Syracuse Region*	*Buffalo-Rochester Region*	*New York City Region (Two Trainings)*	*Totals*
NYSAA Scoring Training	44	26	33	94	197

8.1 SCORING OF OPERATIONAL TESTS

The scoring of NYSAA datafolios occurs during the spring, following the close of the administration period. Scoring is a decentralized process carried out at regional Scoring Institutes. The Department provides a scoring window within which the institutes conduct their scoring sessions. The

purpose of the Scoring Institute is to provide a forum for educators to score the NYSAA student datafolios. Each Scoring Institute is overseen by an SSC and an AATN Specialist. These individuals are thoroughly trained and participate in a qualifying process conducted by the Department and Measured Progress. They are each given a duplicate set of training materials that are to be used during turnkey training at their own Scoring Institutes. They are required to follow the model of the training process demonstrated by the Department and Measured Progress.

There are a variety of processes involved in the Scoring Institute. The basic outline for the review of student datafolios consists of three major steps. Scorers review student datafolios; confirm that the Connection to Grade-Level Content, including the baseline data point performance information, is satisfied; and verify the percentage and rating for Level of Accuracy is documented by the teacher for each Extension or Alternate Grade-Level Indicator (AGLI) assessed. Any questions that arise during scoring are directed to a Table Leader. Scorers use the document titled Steps for Scoring 2014–15 NYSAA Datafolios as the main reference for scoring each datafolio. Table Leaders use the Decision Rules for Scoring 2014–15 NYSAA Datafolios as a reference document for any questions that are not addressed in the Steps for Scoring 2014–15 NYSAA Datafolios. Both documents are included in this report, as Appendices B (Scoring Procedures) and C (Scoring Decision Rules).

On a worksheet, a Scorer records the Extension or AGLI code, Connection to Grade-Level Content questions, percentages for the Level of Accuracy for the baseline administration and final administration, whether or not the student was prompted, and any Scorer comments. Part of this worksheet is returned to the school district along with the datafolio for review by the instructional team and administrators.

Once a datafolio has been reviewed completely, the last step is for the Scorer to transcribe the Extension or AGLI codes, Connection to Grade-Level Content questions, percentages, and other information onto a Scannable Score Document (Scannable). The score document is scanned by the Regional Information Center (RIC) or the Big Five City Scan Center. (The Big Five City School Districts are Buffalo, New York City, Rochester, Syracuse, and Yonkers, each having its own City Scan Center.)

8.2 SCORING RUBRIC

The Scoring Rubric is the initial guide that drives the model used to score NYSAA datafolios. The Scoring Rubric is provided in the 2014–15 *NYSAA Administration Manual*, along with guidance on the process that teachers must follow to meet the scoring requirements. The rubric is broken into two parts. The first part outlines the grades and content requirements, and provides some brief assessment requirements information. The second part provides information about the factors for a performance level. The factors included are the Connection to Grade-Level Content, student performance, and Level of Complexity. The Connection to Grade-Level Content is explained on the Scoring Rubric as follows: “Extensions/AGLIs are assessed based on the appropriate grade level academic content for students

with severe cognitive disabilities. The Assessment Task must align to the Extension/AGLI chosen AND the verifying evidence must be aligned to the task. If these connections are not clear, the Extension/AGLI will not be scored." The final administration Level of Accuracy provides the percentage for the performance dimension. For each Assessment Task documented, the percentage for Level of Accuracy (relative to the student's demonstration of skills, in relation to the Extension or AGLI) and the Level of Complexity that the Extension or AGLI came from combine to give the overall performance level. The Scoring Rubric is presented in Table 8-2 and the Factors for a Performance Level are presented in Table 8-3.

Table 8-2. 2014–15 NYSAA: Scoring Rubric

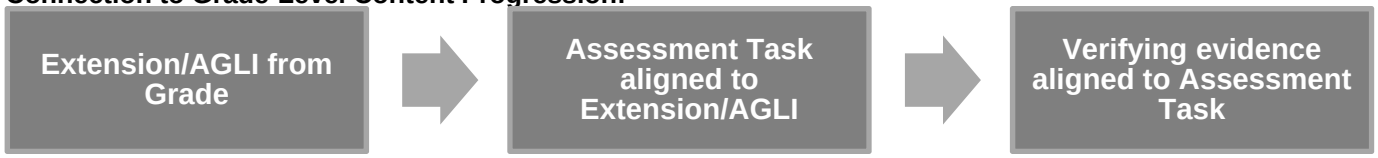
Students with disabilities participating in the NYSAA are assessed according to chronological ages aligned to grade levels. Refer to the Age Range Chart for current date of birth ranges. Students should be tested only once at each grade and in all of the content areas indicated for each grade. For all content areas, student performance data are collected on at least two dates within the administration period. Baseline data must be collected to confirm that the student has not yet mastered the selected Extension or AGLI.

Grade	*ELA*	*Mathematics*	*Science*	*Social Studies*
3	5 Standards	5 Standards		
4	5 Standards	5 Standards	2 Standards	
5	5 Standards	5 Standards		
6	5 Standards	5 Standards		
7	5 Standards	5 Standards		
8	5 Standards	5 Standards	2 Standards	
High School	5 Standards	5 Standards	2 Standards	2 Standards

Table 8-3. 2014–15 NYSAA: Factors for a Performance Level: Connections to Grade-Level Content, Performance, Level of Complexity

Connection to Grade-Level Content = Extensions/AGLIs are assessed based on the appropriate grade level academic content for students with severe cognitive disabilities. The Assessment Task must align to the Extension/AGLI chosen AND the verifying evidence must be aligned to the task. If these connections are not clear, the Extension/AGLI will not be scored.

Connection to Grade-Level Content Progression:



Performance = Level of Accuracy (%)	
Level of Accuracy	The student demonstrates skills based on the Extensions or AGLIs resulting in a percentage for Level of Accuracy.
Independence	Was the student prompted in any way during the administration of the Assessment Task? Yes or No.

Level of Complexity	Less Complex	Middle	More Complex

No or No Score (NS) results when one or more of these issues are identified during scoring (including but not limited to)		
Connection to Grade-Level Content	Performance	Level of Complexity
• Required standard not assessed • Extension or AGLI assessed from incorrect grade • Incorrect Assessment Task assessed • Verifying evidence does not demonstrate task	• Required data points and/or evidence not submitted • Required elements not documented on evidence • Verifying evidence not valid	• Score for baseline administration over threshold (Level of Accuracy is 75% or higher)

8.3 SCORING PROCESS AND RELIABILITY MONITORING REVIEW

8.3.1 Scoring Process

Scorers, who are all New York State teachers or other licensed and/or certified professionals, are directed to objectively review and document the ratings for student performance data contained in the datafolio. During the Scoring Training, it is explained that the data provide an opportunity for students to demonstrate their knowledge, skills, and understanding of the grade-level content. Scoring processes are consistent from one grade level to the next. The same procedures and rules apply to all grade levels and content areas, which is critical to the procedural validity of the assessment.

Scoring Training includes a DVD presentation, a series of guided practice samples, and the Scorer qualification process. (These are described in further detail in the next section.)

The actual scoring process involves reviewing the datafolio compiled by the teacher. The review is meant to ensure that all of the requirements are met. The Scorer records the rubric rating for each Extension or AGLI assessed. If the Connection to Grade-Level Content and the baseline administration performance are satisfied, the final performance percentages can be confirmed, and each performance percentage for baseline and final administrations can be recorded by the Scorer. If the Connection to Grade-Level Content is not met or the baseline administration performance is above the percentage threshold, a rating of No Score (NS) is recorded. After the Scoring Institute, the Scorer ratings are converted to the alternate assessment performance levels, which appear on the NYSAA reports.

In order for Scorers to complete their review of the datafolios, a set of standardized tools is provided to each Scoring Institute. These tools include the *NYSAA Administration Manual* and Frameworks, Scoring Procedures, Scoring Decision Rules, Guided Practices, and qualifier sets. Student performance ratings are documented on a Scorer Worksheet, with a Menu of Comments, and a Scannable. The Menu of Comments, located on the back of each page of the Scorer Worksheet, includes information that a Scorer records when an Extension or AGLI has an NS rating. It also allows the Scorer to provide additional constructive feedback to a teacher about the datafolio.

There are 14 steps involved in the scoring process. The step-by-step procedures outlined in the Steps for Scoring 2014–15 NYSAA Datafolios are implemented statewide and ensure scoring reliability across all Scoring Institutes. Table 8-4 presents a quick review of the steps.

Table 8-4. 2014–15 NYSAA: Scoring Steps Quick Reference

Step(s)		
1		Student demographics, Scorer ID, Scoring Institute code, confirm student's date of birth and grade level assessed, Testing Accommodations, and Collegial Review
2a and b		Review sequence of documentation for content area
3		Demographic information on DSS complete and accurate when compared to the Student Page
4		Extension or AGLI from grade level (Connection to Grade-Level Content)
5		Task connects to Extension or AGLI (Connection to Grade-Level Content)
6a, b, and c		Verifying evidence connects to task (Connection to Grade-Level Content), and Level of Accuracy for the baseline data point is 74% or below
7		Dates on DSS within the administration period
8a-f		Valid verifying evidence and supporting evidence
	a	Valid verifying evidence and supporting evidence: required elements clearly documented (3)
	b	Valid verifying evidence: Student Work Product: Original
	c	Valid verifying evidence: Data Collection Sheet (DCS): Minimum of three dates; includes supporting evidence and staff initials
	d	If verifying evidence is DCS, supporting evidence is present and valid
	e	Valid verifying evidence: Photographs: Minimum of three sequential, captioned, and dated photographs
	f	Valid verifying evidence: digital video or audio clip: Clip is brief and has recorded markers
9		Supports provided that guided the student to the correct answer
10		Confirm final administration percentage for Level of Accuracy, record percentages for Level of Accuracy for final and baseline administrations, record if student was prompted for final and baseline administrations
11		Score the second Extension or AGLI (Steps 3–10)
12		Score mathematics, science, and social studies (Steps 2–11)
13		Confirm Scorer Worksheet is complete, including Procedural Error Comments and additional Scorer Comments
14		Complete the Scannable Score Document

The Scoring Procedures are separated into two major sections: preparing to score, and reviewing and scoring a datafolio. Each step asks the Scorer a question or directs the Scorer to confirm a certain requirement. The steps are presented in a yes/no format to assist the Scorer in moving from one step to another. If a Scorer encounters a "no" or an issue outside of the directions provided in the Scoring Procedures, he or she must consult with the Table Leader. The Table Leader refers to the Decision Rules for Scoring 2014–15 NYSAA Datafolios, if the information on how to proceed in scoring the datafolio is not already provided in the Scoring Procedures.

The Scoring Decision Rules have their own segment in the training DVD. There is also a brief overview of the Decision Rules within the Scoring Procedures segment of the training DVD. The Decision Rules serve as guidance for Table Leaders when a Scorer encounters an issue that is outside of the direction provided in the Scoring Procedures document. The rules are organized by topic, beginning with rules that apply to the datafolio as a whole (e.g., incorrect forms, missing Student Page, evidence of photocopies, correction fluid/tape or black out). The other topic headings are "Assessment Task," "Verifying Evidence," and "Dates." Fifteen Decision Rules were developed that are based on actual datafolio issues found during a Benchmarking review of datafolios in progress. In the training

DVD, Scoring Decision Rules are presented by number as found in the Decision Rules chart. If possible, an example is provided that highlights the point of the Decision Rule, and a description is provided regarding how the rules are to be consistently applied statewide at each Scoring Institute.

8.3.2 Reliability Monitoring Review

The purpose of the Reliability Monitoring Review (RMR) is to ensure scoring consistency and reliability across Scoring Institutes.

At the end of the Scoring Institute, 20% of the scored datafolios from each scoring site are randomly collected by the SSC for the RMR. Measured Progress conducts a Scoring Institute in which the random datafolios are scored by highly experienced and qualified Scorers. RMR Scorers complete the same NYSAA training and qualification process that is used statewide.

RMR scores are compared with the original scores from the regional Scoring Institutes. The original score remains the score of record; the RMR score does not change or affect the original score in any way. The 2014–15 RMR results are presented in Chapter 10.

8.4 Scorer Qualification and Training

A standardized statewide process for Scorer Training and qualification is observed. Each Board of Cooperative Educational Services (BOCES) and Big Five City School District conducts at least one two-day Scoring Institute during the scoring period. For 2014–15, the scoring period was March 16, 2015 to May 1, 2015. The same training and scoring process, Scoring Procedures, and Decision Rules were applied and implemented statewide.

The DVD presentation portion of the training includes a welcome and an introduction, which briefly outline the DVD segments and documents used during training. The DVD then outlines the scoring tools, the step-by-step process for reviewing the datafolios and documenting student scores, and the practice scoring that is done while following along with the DVD segment. The first practice is completed according to directions outlined in the DVD segment. The first Extension is completed as part of the DVD segment. The DVD segment is then paused and participants complete the second Extension. The second Extension is completed as a group or in pairs. The DVD segment provides details about how the second Extension should have been scored.

After the first two DVD segments, Scorers practice scoring two additional datafolio samples with two Extensions or AGLIs—first as a group or in pairs, and then individually. Each practice is reviewed to ensure that Scorers are following the Scoring Procedures accurately. The final DVD segment details other best practice information for scoring, and reinforces information about confirming connection of the verifying evidence to the Assessment Task and Data Collection Sheets. It also provides details about the subsequent steps in Scorer Training.

After the DVD, Scorers are given an opportunity for final questions. Training ends with Scorers completing three calibrated qualifiers with two Extensions or AGLIs each. The qualifiers are actual student datafolios in a content area. The qualifiers were identified by a group of special education teacher and general education teacher committees during a Benchmarking process. Each Scorer must earn a score of 80% or higher to become qualified. Scorers who do not qualify on the first qualifier set receive additional training and must complete an additional qualification sample. After the initial set, Scorers have three opportunities to receive retraining and to qualify. If a Scorer does not qualify after three additional attempts, he or she is reassigned to another role in the Scoring Institute.

8.5 SCORING QUALITY CONTROL

The Quality Control Process at each Scoring Institute is handled by the SSC, AATN Specialists, and Table Leaders. The SSC is responsible for planning, conducting, and coordinating NYSAA scoring activities for the regional Scoring Institute. Each BOCES or Big Five City School District designates at least one individual to assume the role of SSC.

SSC responsibilities include:

- ensuring that the Scoring Procedures, Decision Rules, and other scoring-related guidelines are implemented consistently per the Department's prescribed model;
- ensuring the lock-and-key security of all datafolios during storage and throughout all scoring sessions (datafolio security must be maintained throughout this process);
- gathering the NYSAA student registration information from the RIC or Big Five City Scan Center to assist in planning the Scoring Institute;
- planning, coordinating, and conducting the Scoring Institute for each BOCES or Big Five City School District;
- being present at all times while scoring is in session;
- coordinating the selection of sample datafolios as requested by the Department for RMR;
- ensuring that scoring documentation is completed and provided to the RIC or Big Five City Scan Center;
- collecting feedback regarding the Scoring Institute from AATN Specialists, Table Leaders, and Scorers;
- providing feedback to the Department about the scoring process, procedures, and documentation; and
- returning datafolios following scoring.

AATN Specialists are designated by each BOCES or Big Five City School District to conduct information sessions and NYSAA training and to assist with scoring.

For NYSAA scoring, AATN Specialists:

- assist SSCs in the planning of the Scoring Institute, as needed;
- conduct training sessions and facilitate qualification sessions for Table Leaders and Scorers;
- act as Floor Managers during the scoring process;
- resolve Table Leader questions, using scoring guidelines and resources;
- participate in the Read Behind Process; and
- provide feedback to SSCs and the Department about the scoring processes, procedures, and documentation.

Table Leaders are integral to making sure that the processes and procedures outlined by the Department in the Scoring Training are followed at each scoring station during each Scoring Institute. There is one Table Leader for every five Scorers.

For NYSAA Scoring, Table Leaders must:

- be experienced Scorers familiar with the 2014–15 NYSAA;
- complete Scoring Training, including the qualification process, prior to the start of the Scoring Institute;
- coordinate the datafolio flow at their assigned scoring stations;
- resolve questions from Scorers, using scoring guidelines and resources;
- review and confirm all adjustments and all No Scores documented by Scorers;
- conduct quality control checks of scored datafolios;
- manage the Read Behind Process;
- separate copies of the Scorer Worksheet, as designated by the SSC;
- return scored datafolios to the appropriate boxes;
- provide feedback to the SSC and the Department about the scoring process, procedures, and documentation; and
- assist as needed in evaluating and providing additional training for Scorers who do not qualify during the first round of qualifying.

The Table Leaders are responsible for three main quality control checks. Their first responsibility is to resolve Scorer questions and to confirm NS ratings. When a Scorer questions the Connection to Grade-Level Content or has a question about scoring a datafolio that may result in an NS, the datafolio must be reviewed with the Table Leader. If the issue cannot be readily resolved by the Table Leader by using the Scoring Procedures and Scoring Decision Rules, it must be brought by the Table Leader to the Floor Manager. If the issue cannot be readily resolved by the Floor Manager, the SSC will make the final decision.

The second responsibility of a Table Leader is to complete a standardized quality control check. A quality control check is conducted by the Table Leader once a datafolio has been scored and returned by a Scorer. The Scorer Worksheet is cross-checked against the Scannable. Any corrections made to the ratings by the Scorer are double-checked, and comments are confirmed as being appropriate. A blue dot is affixed by the Table Leader to confirm that the quality control check was conducted.

The third responsibility of a Table Leader is to manage the Read Behind Process. The Read Behind Process occurs throughout the Scoring Institute. This process ensures the integrity of scoring across scoring stations. Table Leaders select the first, third, and then every seventh datafolio from each Scorer for a read behind. The Scannable is pulled and held by the Table Leader, and a red dot is placed on the datafolio. This indicates that it has been selected for a read behind. The first Scorer scores the datafolio, completes the Scorer Worksheet, and returns the datafolio to the Table Leader. The Table Leader turns the Scorer Worksheet over, places it into the front pocket of the datafolio, and then routes the scored datafolio to be scored at a different scoring station or a read-behind table for the second read. The second Scorer scores the datafolio, completes a second Scorer Worksheet, and returns the datafolio to the Table Leader. The Table Leader (either at the first scoring station or read-behind table) compares the two worksheets. If no discrepancy exists, the Table Leader fills in his or her Scorer ID# and completes the Scannable. A quality control check is completed, and a blue dot is affixed to the datafolio. The second Scorer Worksheet is destroyed. If a discrepancy between the scores is found, the Table Leader highlights the discrepant areas and forwards the datafolio to the Floor Manager or SSC for resolution. The Floor Manager or SSC reviews the discrepant areas, enters his or her Scorer ID#, and completes the Scannable. The Floor Manager returns the datafolio to the Table Leader for quality control. After a datafolio has been through the Read Behind Process, the Table Leader completes a quality control check. The Table Leader then works with the Scorer to review the discrepancy and provides any training or support that is needed. If the Scorer continues to have discrepant scores, the Table Leader is then directed to consult the Floor Manager and/or the SSC to discuss additional training or reassignment.

As an additional quality control check to confirm that the Scoring Institutes are following all of the processes and guidelines prescribed by the Department, score site observation visits to a sample of Scoring Institutes are conducted. Each year, the Department designates a set of sites to be monitored during their Scoring Institutes. The observation visits are conducted by the Regional Lead Trainers (RLTs). SSCs are notified if they are selected by the Department for observation. Observers cannot participate or assist in any part of the Scoring Institute. They cannot interact or provide technical assistance during the observation. An Observation Protocol Checklist is completed during the visit and is then submitted to the Department.

CHAPTER 9 CLASSICAL ITEM ANALYSIS

As noted in Brown (1983), “A test is only as good as the items it contains.” A complete evaluation of a test’s quality must include an evaluation of each item. Both the *Standards for Educational and Psychological Testing* (AERA et al., 2014) and the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) include standards for identifying quality items. While the specific statistical criteria identified in these publications were developed primarily for general—not alternate—assessments, the principles and some of the techniques apply within the alternate assessment framework, as well.

Both qualitative and quantitative analyses were conducted to ensure that New York State Alternate Assessment (NYSAA) items met these standards. Qualitative analyses are described in earlier sections of this report; this section focuses on the quantitative evaluations. The statistical evaluations discussed are difficulty indices, discrimination (item-test correlations), item means, structural relationships (correlations between the dimensions), and bias and fairness. The item analyses presented here are based on the statewide administration of the 2014–15 NYSAA.

9.1 DIFFICULTY AND DISCRIMINATION

For the NYSAA, each student datafolio for a specified content area at a given grade level receives an Accuracy score on each of five standards in English Language Arts (ELA) and mathematics, and on each of two standards in science and social studies. For each standard, teachers choose a task by which to assess their students. The chosen task may be at one of three possible Levels of Complexity (LOC: LOC1, LOC2, or LOC3), where the higher levels indicate greater complexity. For a given student, the LOC at which the student is assessed may differ from one standard to another. Thus, for any one standard, the number of students assessed varies across the LOCs, and the way that the student counts vary across the LOCs varies across the standards. Tables H-1 to H-18 in Appendix H, Classical Item Analysis, include the student counts for the three LOCs for each assessed standard. Table 9-1 summarizes the means and ranges of these counts. As can be seen in these tables, approximately 3,000 students were assessed at each grade level; on average, about 45% of the students were assessed at LOC1, about 42% at LOC2, and about 14% at LOC3. Note, however, that there also existed substantial variability across standards in how students were distributed across the LOCs. For example, for Grade 3 ELA, for Standard 314 about 75% of the students were assessed at LOC2, whereas for Standard 322 about 72% were assessed at LOC1. In general, for every grade level and content area, LOC1 and LOC2 combined to easily have the most assessed students. The largest count of students assessed at LOC3 occurred for Grade 7 ELA Standard 753, where the 1,278 students were only 38% of the total.

Table 9-1. 2014–15 NYSAA: Summary of Numbers of Students Assessed Across the Levels of Complexity

Subject	*Grade*	*LOC 1*			*LOC 2*			*LOC 3*		
		Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
English Language Arts	3	1170.60	374	1,942	1178.20	601	2,022	366.60	158	505
	4	1260.60	500	2,002	1395.60	573	2,108	272.80	112	437
	5	1248.20	790	1,854	1662.80	1,014	2,200	169.80	77	246
	6	1234.20	702	1,609	1680.80	1,369	2,381	249.60	76	556
	7	1932.00	1,231	2,387	850.80	289	1,869	523.80	157	1,278
	8	2000.20	1,694	2,237	980.40	626	1,249	351.00	242	459
	High School	1240.20	749	2,013	1186.20	428	1,587	420.80	324	519
Mathematics	3	1115.00	590	1,756	1257.20	542	1,821	346.40	92	587
	4	1528.00	895	2,055	895.60	316	1,383	506.40	84	1,031
	5	1687.40	748	2,449	1042.20	440	2,088	348.60	155	777
	6	1505.20	1,227	1,845	1333.20	646	1,782	322.60	108	680
	7	1548.80	537	2,557	1153.00	317	2,353	609.80	181	1,165
	8	2168.20	1,866	2,821	810.00	376	1,173	359.40	163	763
	High School	1142.40	643	1,494	1009.60	478	1,889	677.60	276	860
Science	4	1347.00	1,289	1,405	1257.00	1,205	1,309	308.00	292	324
	8	1079.00	743	1,415	1755.00	1,320	2,190	498.00	404	592
	High School	856.00	654	1,058	1343.00	777	1,909	657.00	294	1,020
Social Studies	High School	632.50	528	737	1793.00	1,662	1,924	422.50	388	457

For each task, the teacher assessed each student a certain number of times (determined by the teacher), and the teacher recorded the number of times the student was successful. Because the number of times a student was assessed for each task was something teachers could decide for themselves, this number varied across students. Therefore, a percent-correct score was recorded rather than a number-correct score. Hence, for a standard assessed at a particular LOC, the observed student scores ranged from 0 to 100.

To develop a single scale for comparing all students for a specific standard, percent-correct task scores associated with higher LOCs deserve more credit than the same scores at lower LOCs. Based on a scientific study jointly carried out by Measured Progress and the Department, a single scale was developed that allows scores at different LOCs for a given standard to be combined into a single scaled score for that standard. Specifically, the scaled score for a given standard is calculated by taking the observed percent-correct score, adding a credit when the LOC is a 2 or 3, and then dividing the result by 10. The complexity credit was 75 for an LOC of 2, and 150 for an LOC of 3. The division by 10 was needed to eliminate unwanted gaps in the scale, resulting from some combinations of scores being much less likely than others. Thus, the scaled scores for a given standard could range from 0 to 25. The formulas for the individual standard scaled scores by complexity are as follows:

$$SS_{LOC1} = RS_{LOC1}/10$$
$$SS_{LOC2} = (RS_{LOC2} + 75)/10$$
$$SS_{LOC3} = (RS_{LOC3} + 150)/10$$

To compare students at the level of total test score, a scaled score that is simply the sum of the scaled scores on the standards plus an additive constant is produced. Specifically, for ELA and mathematics, the total score is the sum of the scaled scores on the five standards plus 400, resulting in total scores that can range from 400 to 525. And for science and social studies, the total score is the sum of the scaled scores on the two standards plus 550, resulting in total scores that can range from 550 to 600. Thus, the formulas for the total subject standard scaled scores are as follows:

$$SS_{Total\ ELA/Math} = SS_{Std1} + SS_{Std2} + SS_{Std3} + SS_{Std4} + SS_{Std5} + 400$$
$$SS_{Total\ Sci/SS} = SS_{Std1} + SS_{Std2} + 550$$

From the above, it is clear that there are two types of scores on this test that could be treated as traditional "item" scores for the purposes of psychometric evaluation. The standards are one reasonable choice to represent the traditional items on the test because each student is assessed on the same number of standards, and the sum of the scaled scores that he or she receives on the standards provides the basis for the total scaled score for a student. Alternatively, the application of a standard at a given LOC (S/LOC) could also be used to represent a traditional item because the raw percent-correct scores are directly comparable across students who were assessed on the same standard at the same LOC.

Using both of these item representations ("Standard" and "S/LOC"), all items were evaluated in terms of item difficulty according to standard classical test theory practices. "Difficulty" was defined as the average proportion of points achieved on an item, and was measured by obtaining the average score on an item and dividing by the maximum score for the item. By computing the difficulty index as the average proportion of points achieved, the items are placed on a scale that ranges from 0.0 to 1.0. Although the p-value is traditionally described as a measure of difficulty (as it is described here), it is properly interpreted as an easiness index, because larger values indicate easier items.

An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item. Items that have either a very high or a very low difficulty index are considered to be potentially problematic because they are either so difficult that few students get full credit or so easy that nearly all students get full credit. In either case, such items should be reviewed for appropriateness for inclusion on the assessment.

It is worth mentioning that using a norm-referenced criterion such as p-values to evaluate test items is somewhat contradictory to the purpose of a criterion-referenced assessment like the NYSAA. Criterion-referenced assessments are intended primarily to provide evidence of student progress

relative to a standard, rather than to differentiate between students. Thus, the generally accepted criteria regarding classical item statistics are only cautiously applicable to the NYSAA.

A desirable feature of an item is that higher-ability students perform better on the item than do lower-ability students. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of an item. Within classical test theory, this item-test correlation is referred to as the item's "discrimination" because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. The discrimination index used to evaluate NYSAA items was the Pearson product-moment correlation. The theoretical range of this statistic is -1.0 to 1.0.

Discrimination indices can be thought of as measures of how closely an item assesses the same knowledge and skills assessed by other items contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. For the NYSAA, the total test scaled score, excluding the item being evaluated, was used as the criterion score.

A summary of the item difficulty and item discrimination statistics for each grade/content area combination is presented in Tables 9-2 and 9-3 for the two kinds of items discussed above, "S/LOC" and "Standard," respectively. As shown in Table 9-2, the mean difficulty values for S/LOC items ranged from 0.83 to 0.89, indicating that, overall, students performed well on the S/LOC items on the NYSAA, and that students, on average, were well prepared for the LOCs that were targeted by their instruction for each standard. On the other hand, as shown in Table 9-3, the mean difficulty values for the standards-based scaled item scores ranged from 0.47 to 0.62, with standard deviations (across the five standards) on the order of about 0.05. These results indicate that the difficulty levels of the five standards within a given assessment are similar to each other (small standard deviation) and are well aligned with the proficiency distributions of the students (means close to 0.50); neither overly difficult nor overly easy. In contrast to alternate assessments, the difficulty values for assessments designed for the general population tend to be in the 0.4 to 0.7 range for the majority of items. Because the nature of alternate assessments is different from that of general assessments, and because very few guidelines exist as to criteria for interpreting these values for alternate assessments, the values presented in Tables 9-2 and 9-3 should not be interpreted to mean that the students who took the NYSAA performed either better or worse than the students who took general assessments.

Also shown in Tables 9-2 and 9-3 are the mean discrimination values for the S/LOC and standards-based items, respectively, as calculated by the correlation between the item scores and the scaled total scores. Because the majority of students received high scores on the S/LOC items and these raw scores are not adjusted for LOC, as is done for the scaled total scores, the discrimination indices are somewhat lower than one might otherwise expect, with mean values ranging from 0.29 to

0.52. In particular, if all of the students received high percent-correct scores on the S/LOC items, there is little variability in the item score for differentiating the criterion scores. On the other hand, for the standards-based items, the mean discrimination indices ranged from 0.61 to 0.76 for mathematics and ELA and 0.52 to 0.64 for science and social studies. These results indicate a strong positive relationship between the scaled scores on the standards-based items and the scaled total scores. As with the item difficulty values, because the nature and use of the NYSAA are different from those of a general assessment, and because very few guidelines exist as to criteria for interpreting these values for alternate assessments, the statistics presented in Tables 9-2 and 9-3 should be interpreted with caution.

Table 9-2. 2014–15 NYSAA: Summary of Item Difficulty and Discrimination Statistics by Content Area and Grade for S/LOC Items

Content Area	*Grade*	*Number of Items*	*Difficulty (p-Value)*		*Discrimination*	
			Mean	*SD*	*Mean*	*SD*
English Language Arts	3	15	0.86	0.07	0.47	0.16
	4	15	0.87	0.06	0.39	0.17
	5	15	0.87	0.05	0.42	0.14
	6	15	0.89	0.06	0.42	0.14
	7	15	0.88	0.04	0.38	0.13
	8	15	0.88	0.05	0.42	0.11
	High School	15	0.85	0.06	0.46	0.13
Mathematics	3	15	0.84	0.07	0.52	0.15
	4	15	0.85	0.06	0.44	0.13
	5	15	0.85	0.05	0.46	0.16
	6	15	0.87	0.06	0.48	0.13
	7	15	0.85	0.07	0.47	0.10
	8	15	0.87	0.06	0.45	0.13
	High School	15	0.83	0.07	0.46	0.10
Science	4	6	0.89	0.06	0.29	0.14
	8	6	0.87	0.06	0.35	0.11
	High School	6	0.86	0.06	0.37	0.10
Social Studies	High School	6	0.84	0.11	0.39	0.09

Table 9-3. 2014–15 NYSAA: Summary of Item Difficulty and Discrimination Statistics by Content Area and Grade for Standards-Based Items

Content Area	*Grade*	*Number of Items*	*Difficulty*				*Discrimination*	
			Scaled Score		p-*Value*		*(Corr. w/Total)*	
			Mean	*Std*	*Mean*	*Std*	*Mean*	*Std*
English Language Arts	3	5	13.71	2.17	0.55	0.09	0.70	0.04
	4	5	13.51	1.67	0.54	0.07	0.63	0.04
	5	5	13.39	1.10	0.54	0.04	0.61	0.02
	6	5	13.89	0.71	0.56	0.03	0.68	0.02
	7	5	12.92	1.45	0.52	0.06	0.65	0.02
	8	5	12.33	0.48	0.49	0.02	0.71	0.03
	High School	5	13.76	1.34	0.55	0.05	0.75	0.03

continued

Content Area	Grade	Number of Items	Difficulty				Discrimination	
			Scaled Score		p-Value		(Corr. w/Total)	
			Mean	Std	Mean	Std	Mean	Std
Mathematics	3	5	13.69	1.67	0.55	0.07	0.72	0.03
	4	5	13.16	1.68	0.53	0.07	0.70	0.04
	5	5	12.63	2.03	0.51	0.08	0.68	0.05
	6	5	13.15	0.65	0.53	0.03	0.76	0.04
	7	5	13.64	2.06	0.55	0.08	0.64	0.03
	8	5	11.79	1.38	0.47	0.06	0.72	0.03
	High School	5	14.46	0.31	0.58	0.01	0.68	0.03
Science	4	2	13.56	0.48	0.54	0.02	0.55	0.01
	8	2	14.77	0.72	0.59	0.03	0.52	0.01
	High School	2	15.58	0.59	0.62	0.02	0.62	0.00
Social Studies	High School	2	15.39	0.13	0.62	0.01	0.64	0.02

9.2 STRUCTURAL RELATIONSHIP

By design, the performance level classification of the NYSAA is based on scaled scores associated with five standards for ELA and mathematics and on two standards for science and social studies. These different standards can be conceptualized as different construct dimensions of the assessment. As with any assessment, it is important that the dimensions composing the assessment be carefully examined. This was achieved by exploring the relationships between student scaled scores on the different dimensions with Pearson correlation coefficients. A very low correlation (near zero) would indicate that the dimensions are not related; a low negative correlation (approaching -1.00) would indicate that they are inversely related (i.e., that a student with a high score on one dimension had a low score on the other); and a high positive correlation (approaching 1.00) would indicate that the information provided by one dimension is similar to that provided by the other dimension. In addition, the correlation matrices for the standards were analyzed with factor analysis to determine the number of dimensions that are statistically significant for each assessment instrument analyzed. Because these assessments are unidimensionally scored, it is important to determine the degree to which unidimensionality accounts for the variability in the scores.

The average correlations between the scaled scores on the standards by content area and grade are shown in Table 9-4. The detailed results for each pair of standards are given in Appendix I.

Table 9-4. 2014–15 NYSAA: Average Correlations

Content Area	Grade	Average Correlation	Correlation Standard Deviation
English Language Arts	3	0.58	0.06
	4	0.51	0.05
	5	0.49	0.04
	6	0.57	0.02
	7	0.53	0.03

continued

Content Area	*Grade*	*Average Correlation*	*Correlation Standard Deviation*
English Language Arts	8	0.61	0.03
	High School	0.66	0.04
Mathematics	3	0.61	0.04
	4	0.59	0.05
	5	0.57	0.06
	6	0.67	0.04
	7	0.52	0.04
	8	0.62	0.03
	High School	0.57	0.05
Science	4	0.59	
	8	0.55	
	High School	0.65	
Social Studies	High School	0.67	

The inter-item correlations ranged from a low of .44 (e.g., Grade 4, Standards 413 and 432) to a high of 0.73 (e.g., Grade 6 mathematics, Standards 606 and 608). The averages for a given grade and content area ranged from 0.49 to 0.67. These correlations indicate that the scores on the different standards within a grade and content area have strong positive relationships with each other. Next, a factor analysis was conducted on the correlation matrix for each grade and content area, to determine the degree to which a unidimensional scale can account for the variance in the scaled total scores. For science and social studies, because they have only two scored dimensions, the strong positive correlations are alone strong evidence in support of their unidimensional scales. As shown in Table 9-5, the factor analysis confirmed this by indicating that 77% to 84% of the variance in the correlations of the standards-based scaled scores is accounted for by a unidimensional scale. For ELA and mathematics, the factor analysis results indicate that a single scored dimension accounts for 59% to 74% of the variance in the correlations of the standards-based scaled scores.

Table 9-5. 2014–15 NYSAA: Factor Analysis Table

Grade	*Content Area*	*Percent of Variance Accounted for by Each Factor*				
		1	*2*	*3*	*4*	*5*
3	English Language Arts	66.5%	11.3%	8.5%	7.6%	6.1%
	Mathematics	68.9%	9.8%	8.2%	7.0%	6.2%
4	English Language Arts	60.5%	12.1%	10.9%	8.8%	7.7%
	Mathematics	67.3%	9.9%	9.1%	7.1%	6.6%
	Science	79.5%	20.5%			
5	English Language Arts	59.1%	12.4%	10.4%	9.9%	8.3%
	Mathematics	65.6%	11.4%	8.8%	7.7%	6.5%
6	English Language Arts	65.5%	9.7%	8.8%	8.3%	7.6%
	Mathematics	74.0%	8.1%	6.7%	5.8%	5.4%
7	English Language Arts	62.4%	11.2%	9.6%	8.5%	8.3%
	Mathematics	62.1%	11.0%	10.2%	8.5%	8.2%

continued

Grade	Content Area	Percent of Variance Accounted for by Each Factor				
		1	*2*	*3*	*4*	*5*
8	English Language Arts	69.3%	9.1%	7.6%	7.6%	6.3%
	Mathematics	69.4%	9.2%	7.9%	7.0%	6.5%
	Science	77.3%	22.7%			
High School	English Language Arts	72.7%	8.6%	7.2%	6.2%	5.2%
	Mathematics	65.8%	10.3%	9.8%	7.2%	6.9%
	Science	82.5%	17.5%			
	Social Studies	83.6%	16.4%			

Thus, these results also give strong support for the appropriateness of the use of a unidimensional scale for these assessment instruments.

9.3 BIAS/FAIRNESS

The *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004) explicitly states that subgroup differences in performance should be examined when sample sizes permit, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than construct-irrelevant, factors. The guidelines in the *Code of Fair Testing Practices in Education* are consistent with the relevant sections of the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

When appropriate, the standardization differential item functioning (DIF) procedure (Dorans & Kulick, 1986) is used to identify items for which subgroups of interest perform differently, beyond the effect of differences in overall achievement. However, because the NYSAA uses a datafolio that does not include standard items that are taken by all students, it was not possible to conduct DIF analyses.

Although it is not possible to run quantitative analyses of item bias for the NYSAA, due to data limitations, fairness can be addressed through the assessment Blueprints, which are designed to reflect the core curriculum, as described in detail earlier in this report. The development of the assessment Blueprints, which reflect recommendations laid out in the *Standards for Educational and Psychological Testing*, were designed to ensure that the test is free of any insensitive or offensive material, as well as to ensure alignment with general education grade-level content and to promote higher expectations for students taking the NYSAA.

Issues of fairness are also addressed in the NYSAA Administration and Scoring Procedures. Chapter 7 of this report describes in detail the procedures for administering the NYSAA and constructing the datafolio, as well as the training and review steps designed to ensure that the test is administered appropriately and consistently to all students. Chapter 8 describes, in detail, the Scoring Rubrics used, selection and training of Scorers, and scoring Quality Control Processes. These processes were followed to minimize bias due to differences in how individual Scorers award scores.

CHAPTER 10 CHARACTERIZING ERRORS ASSOCIATED WITH TEST SCORES

One of the primary uses of the New York State Alternate Assessment (NYSAA) scores is for school-, district-, and State-level accountability in the federal No Child Left Behind (NCLB) Act 2001 and in State accountability systems. The students are classified as Proficient or Not Proficient, and are included in the State's Adequate Yearly Progress (AYP) calculation. In this case, the reliability of individual student scores, while not meaningless, becomes much less important. The scores have been collapsed for each student to a yes/no decision and then aggregated across students.

For purposes of calculating reliability estimates, standards-based item scores are defined in the same way as described in Chapter 9. Specifically, the scaled scores on the five standards for English Language Arts (ELA) and mathematics and the two standards for science and social studies are treated as the item scores.

10.1 RELIABILITY

In the previous chapter, individual item characteristics of the 2014–15 NYSAA were presented. Although individual item performance is an important focus for evaluation, a complete evaluation of an assessment must address the way in which the items (or, in this case, standards-based items) that make up the test score function together and complement one another. Any measurement includes some amount of measurement error. No academic assessment can measure student performance with perfect accuracy; some students will receive scores that underestimate their true abilities, and other students will receive scores that overestimate their true abilities. Items that function well together produce assessments that have less measurement error (i.e., the error is small, on average). Such assessments are described as reliable.

There are a number of ways to estimate an assessment's reliability. One approach is to split all test items into two groups and then correlate students' scores on the two half-tests. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, the items included are likely measuring very similar knowledge or skills. It suggests that measurement error will be minimal.

The split-half method requires psychometricians to select items that contribute to each half-test score. This decision may have an effect on the resulting correlation, since each different possible split of the test halves will result in a different correlation. Another problem with the split-half method of calculating reliability is that it underestimates reliability, because test length is cut in half. All else being equal, a shorter test is less reliable than a longer test. Cronbach (1951) provided a statistic, alpha (α), that avoids the shortcomings of the split-half method by comparing individual item variances to total test variance. Cronbach's α was used to assess the reliability of the 2014–15 NYSAA tests. The formula is as follows:

$$\propto \equiv \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma^2_{(Y_i)}}{\sigma^2_x}\right],$$

where
i indexes the item,
n is the number of items,
$\sigma^2_{(Y_i)}$ represents individual item variance, and
σ^2_x represents the total test variance.

If the correlation is high (in practice, toward the high end of the typical Cronbach's α range of 0.50 to 0.99), the parts of the test are likely measuring very similar knowledge or skills. Thus, a high Cronbach's α coefficient is evidence that the standards-based items complement one another, and suggests that the assessment is reliable. Table 10-1 presents scaled total score descriptive statistics (maximum possible scaled total score, average scaled total score, and standard deviation), Cronbach's α coefficient, and scaled total score standard errors of measurements (SEMs) for each content area and grade. The results show that the reliability estimates range from 0.82 to 0.91 for ELA and mathematics, and range from 0.70 to 0.80 for science and social studies. The latter values are expected to be lower because those tests have fewer items. Considering that the NYSAAs are necessarily shorter than general assessments, the reliability coefficients in Table 10-1 give strong support to the reliability of the reported scaled total scores and their intended interpretations.

Table 10-1. 2014–15 NYSAA: Scaled Total Score Descriptive Statistics, Cronbach's Alpha, and Standard Errors of Measurement (SEM) by Grade and Content Area

Content Area	*Grade*	*Number of Students*	*Scaled Total Score*			*Reliability (α)*	*SEM*
			Maximum	*Mean*	*Standard Deviation*		
English Language Arts	3	2,745	525	467.50	24.61	0.87	8.79
	4	2,967	525	466.36	21.61	0.83	8.85
	5	3,127	525	465.62	20.03	0.82	8.40
	6	3,212	525	468.10	22.30	0.86	8.23
	7	3,361	525	463.28	24.72	0.84	9.87
	8	3,398	525	460.14	24.97	0.89	8.39
	High School	2,900	525	467.22	26.20	0.91	8.05
Mathematics	3	2,746	525	467.41	24.70	0.88	8.45
	4	2,969	525	464.60	26.27	0.87	9.52
	5	3,129	525	461.82	23.45	0.86	8.67
	6	3,213	525	464.37	25.50	0.90	7.86
	7	3,364	525	466.82	25.47	0.84	10.19
	8	3,392	525	457.68	25.00	0.89	8.46
	High School	2,903	525	470.14	28.28	0.87	10.28
Science	4	2,953	600	576.50	10.54	0.74	5.38
	8	3,382	600	578.83	10.48	0.70	5.75
	High School	2,896	600	580.43	11.59	0.77	5.59
Social Studies	High School	2,886	600	580.17	10.99	0.80	4.89

10.2 SUBGROUP RELIABILITY

The reliability coefficients discussed in the previous section were based on the overall population of students who took the 2014–15 NYSAA. Subgroup Cronbach's α's were calculated using the formula defined above, using only the members of the subgroup in question in the computations. These statistics are reported in Appendix D. Note that statistics are reported only for subgroups with at least 11 students.

For several reasons, the results of this section should be interpreted with caution. First, inherent differences between grades and content areas preclude making valid inferences about the quality of a test that are based on statistical comparisons with other tests. Second, reliabilities are dependent not only on the measurement properties of a test but on the statistical distribution of the studied subgroup. For example, in Appendix D, it can be readily seen that subgroup sample sizes may vary considerably, which results in natural variation in reliability coefficients. Alternatively, α, which is a type of correlation coefficient, may be artificially depressed for subgroups with little variability (Draper & Smith, 1998). Additionally, there is no industry standard to interpret the strength of a reliability coefficient, and this is particularly true when the population of interest is a single subgroup.

10.3 DECISION ACCURACY AND CONSISTENCY

While related to reliability, the accuracy and consistency of classifying students into performance categories is an even more important issue in a standards-based reporting framework (Livingston & Lewis, 1995). Unlike generalizability coefficients, decision accuracy and consistency (DAC) can usually be computed with the data currently available for most alternate assessments. Based on the raw scale cut scores established for each content area via standard setting in June 2008, each student was classified into one of the following performance levels: Not Meeting Learning Standards, Partially Meeting Learning Standards, Meeting Learning Standards, and Meeting Learning Standards with Distinction. (Lookup tables for converting raw scores to performance levels are presented in Chapter 11.) This section of the report explains the methodologies used to assess the reliability of classification decisions and presents the results.

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated, because errorless test scores do not exist. Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete and parallel forms of the test are given to the same group of students. In operational test programs, however, such a design is usually impractical. Instead, techniques have been developed to estimate both the accuracy and the consistency of classification decisions based on

a single administration of a test. The Livingston and Lewis (1995) method was used for the NYSAA because it is easily adaptable to all types of testing formats.

The accuracy and consistency estimates reported in the following tables make use of "true scores" in the classical test theory sense. A true score is the score that would be obtained if a test had no measurement error. Of course, true scores cannot be observed and, therefore, must be estimated. In the Livingston and Lewis method, estimated true scores are used to categorize students into their "true" classifications.

For the NYSAA, after various technical adjustments (described in Livingston & Lewis, 1995), a four-by-four contingency table of accuracy was created for each content area and grade, where cell [i, j] represented the estimated proportion of students whose true score fell into classification i (where i = 1 to 4) and observed score into classification j (where j = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students whose true and observed classifications matched) signified overall accuracy.

To calculate consistency, true scores were used to estimate the joint distribution of classifications on two independent, parallel test forms. Following statistical adjustments per Livingston and Lewis (1995), a new four-by-four contingency table was created for each content area and grade, and was populated by the proportion of students who would be categorized into each combination of classifications according to the two (hypothetical) parallel test forms. Cell [i, j] of this table represented the estimated proportion of students whose observed score on the first form would fall into classification i (where i = 1 to 4) and whose observed score on the second form would fall into classification j (where j = 1 to 4). The sum of the diagonal entries (i.e., the proportion of students categorized by the two forms into exactly the same classification) signified overall consistency.

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classifications that would be expected by chance. It is calculated using the following formula:

$$\kappa = \frac{(\text{Observed agreement}) - (\text{Chance agreement})}{1 - (\text{Chance agreement})} = \frac{\sum_i C_{ii} - \sum_i C_{i.} C_{.i}}{1 - \sum_i C_{i.} C_{.i}},$$

where

$C_{i.}$ is the proportion of students whose observed performance level would be Level i (where i = 1 – 4) on the first hypothetical parallel form of the test,

$C_{.i}$ is the proportion of students whose observed performance level would be Level i (where i = 1 – 4) on the second hypothetical parallel form of the test, and

C_{ii} is the proportion of students whose observed performance level would be Level i (where i = 1 – 4) on both hypothetical parallel forms of the test.

Because κ is corrected for chance, its values are lower than those of other consistency estimates.

The accuracy and consistency analyses described above are provided in Table 10-2. The table includes overall accuracy and consistency indices, including kappa. Accuracy and consistency values conditional upon performance level are also given. For these calculations, the denominator is the

proportion of students associated with a given performance level. For example, the conditional accuracy value is 0.83 for Meeting Learning Standards for Grade 3 ELA. This figure indicates that among the students whose true scores placed them in this classification, 83% would be expected to be in this classification when categorized according to their observed scores. Similarly, a consistency value of 0.79 indicates that 79% of students with observed scores in the Meeting Learning Standards level would be expected to score in this classification again if a second, parallel test form were used.

Table 10-2. 2014–15 NYSAA: Summary of Decision Accuracy (and Consistency) Results by Content Area and Grade—Overall and Conditional on Performance Level

Content Area	*Grade*	*Overall*	*Kappa*	*Conditional on Level*			
				Not Meeting	*Partially Meeting*	*Meeting*	*Meeting with Distinction*
English Language Arts	3	0.80 (0.73)	0.55	0.72 (0.55)	0.68 (0.58)	0.83 (0.79)	0.86 (0.72)
	4	0.82 (0.75)	0.51	0.64 (0.37)	0.69 (0.57)	0.85 (0.82)	0.83 (0.66)
	5	0.81 (0.74)	0.49	0.65 (0.37)	0.66 (0.51)	0.84 (0.81)	0.83 (0.67)
	6	0.82 (0.75)	0.54	0.69 (0.47)	0.69 (0.57)	0.85 (0.82)	0.85 (0.70)
	7	0.77 (0.69)	0.49	0.72 (0.57)	0.49 (0.38)	0.81 (0.78)	0.85 (0.70)
	8	0.78 (0.71)	0.55	0.78 (0.68)	0.47 (0.36)	0.82 (0.77)	0.87 (0.77)
	High School	0.82 (0.75)	0.61	0.77 (0.65)	0.71 (0.61)	0.84 (0.80)	0.88 (0.78)
Mathematics	3	0.80 (0.73)	0.54	0.75 (0.62)	0.61 (0.50)	0.84 (0.81)	0.86 (0.72)
	4	0.78 (0.71)	0.53	0.74 (0.61)	0.57 (0.46)	0.82 (0.77)	0.87 (0.74)
	5	0.81 (0.74)	0.53	0.70 (0.52)	0.63 (0.51)	0.84 (0.81)	0.85 (0.70)
	6	0.82 (0.75)	0.60	0.74 (0.59)	0.72 (0.63)	0.83 (0.79)	0.88 (0.78)
	7	0.78 (0.70)	0.49	0.68 (0.49)	0.63 (0.51)	0.81 (0.78)	0.84 (0.68)
	8	0.79 (0.71)	0.55	0.76 (0.65)	0.62 (0.51)	0.83 (0.78)	0.86 (0.73)
	High School	0.79 (0.72)	0.53	0.71 (0.56)	0.61 (0.49)	0.82 (0.79)	0.87 (0.73)
Science	4	0.76 (0.68)	0.36	0.59 (0.33)	0.51 (0.38)	0.81 (0.79)	0.81 (0.55)
	8	0.76 (0.67)	0.34	0.56 (0.26)	0.54 (0.40)	0.79 (0.77)	0.82 (0.54)
	High School	0.75 (0.67)	0.41	0.59 (0.34)	0.55 (0.42)	0.77 (0.75)	0.85 (0.64)
Social Studies	High School	0.76 (0.67)	0.48	0.63 (0.40)	0.62 (0.50)	0.71 (0.66)	0.88 (0.77)

For some testing situations, the greatest concern may be decisions around level thresholds. For example, in testing done for No Child Left Behind (NCLB) Act of 2001 accountability purposes, the primary concern is distinguishing between students who are proficient and those who are not yet proficient. In this case, the accuracy of the Partially Meeting/Meeting threshold is of greatest interest. Table 10-3 provides accuracy and consistency estimates at each cutpoint, as well as false positive and false negative decision rates. (A false positive is the proportion of students whose observed scores were above the cut and whose true scores were below the cut. A false negative is the proportion of students whose observed scores were below the cut and whose true scores were above the cut.)

The indices described above are derived from Livingston and Lewis's (1995) method of estimating the accuracy and consistency of classifications. It should be noted that Livingston and Lewis discuss two versions of the accuracy and consistency tables. A standard version performs calculations

for forms parallel to the form taken. An “adjusted” version adjusts the results of one form to match the observed score distribution obtained in the data. The tables on the previous pages use the standard version for two reasons: (1) This “unadjusted” version can be considered a smoothing of the data, thereby decreasing the variability of the results; and (2) for results dealing with the consistency of two parallel forms, the unadjusted tables are symmetrical, indicating that the two parallel forms have the same statistical properties. This second reason is consistent with the notion of forms that are parallel; that is, it is more intuitive and interpretable for two parallel forms to have the same statistical distribution.

Note that, as with other methods of evaluating reliability, DAC statistics calculated based on small groups can be expected to be lower than those calculated based on larger groups. For this reason, the values presented in the following tables should be interpreted with caution. In addition, it is important to remember that it is inappropriate to compare DAC statistics between grades and content areas.

Table 10-3. 2014–15 NYSAA: Summary of Decision (and Consistency) Results by Content Area and Grade—Conditional on Cutpoint

Content Area	*Grade*	*Not Meeting / Partially Meeting*			*Partially Meeting / Meeting*			*Meeting / Meeting with Distinction*		
		Accuracy (consistency)	*False*		*Accuracy (consistency)*	*False*		*Accuracy (consistency)*	*False*	
			Positive	*Negative*		*Positive*	*Negative*		*Positive*	*Negative*
English Language Arts	3	0.97 (0.96)	0.01	0.02	0.91 (0.88)	0.04	0.04	0.91 (0.88)	0.06	0.03
	4	0.99 (0.99)	0.00	0.01	0.92 (0.89)	0.04	0.04	0.91 (0.88)	0.07	0.02
	5	0.99 (0.98)	0.00	0.01	0.93 (0.90)	0.03	0.04	0.90 (0.86)	0.07	0.03
	6	0.98 (0.98)	0.00	0.01	0.92 (0.89)	0.04	0.04	0.91 (0.88)	0.06	0.02
	7	0.95 (0.93)	0.02	0.03	0.91 (0.87)	0.04	0.05	0.90 (0.87)	0.07	0.03
	8	0.95 (0.92)	0.02	0.03	0.92 (0.88)	0.04	0.04	0.91 (0.88)	0.06	0.03
	High School	0.97 (0.96)	0.01	0.02	0.93 (0.90)	0.04	0.04	0.92 (0.90)	0.05	0.02
Mathematics	3	0.96 (0.94)	0.02	0.02	0.92 (0.88)	0.04	0.04	0.92 (0.89)	0.05	0.02
	4	0.96 (0.94)	0.02	0.02	0.91 (0.88)	0.04	0.04	0.91 (0.88)	0.06	0.03
	5	0.97 (0.96)	0.01	0.02	0.92 (0.89)	0.04	0.04	0.91 (0.88)	0.06	0.02
	6	0.98 (0.97)	0.01	0.01	0.92 (0.89)	0.04	0.04	0.92 (0.89)	0.05	0.03
	7	0.97 (0.96)	0.01	0.02	0.91 (0.87)	0.05	0.05	0.90 (0.87)	0.07	0.02
	8	0.95 (0.93)	0.02	0.03	0.91 (0.87)	0.05	0.04	0.93 (0.90)	0.05	0.02
	High School	0.97 (0.95)	0.01	0.02	0.92 (0.89)	0.04	0.04	0.91 (0.87)	0.07	0.03
Science	4	0.97 (0.96)	0.01	0.02	0.90 (0.86)	0.05	0.06	0.89 (0.85)	0.09	0.02
	8	0.98 (0.97)	0.00	0.02	0.90 (0.86)	0.05	0.06	0.87 (0.83)	0.11	0.02
	High School	0.98 (0.97)	0.01	0.02	0.91 (0.88)	0.04	0.05	0.86 (0.82)	0.11	0.03
Social Studies	High School	0.98 (0.97)	0.01	0.01	0.91 (0.88)	0.04	0.04	0.86 (0.81)	0.10	0.04

10.4 INTERRATER CONSISTENCY

Chapter 8 of this report describes in detail the processes that were implemented to monitor the quality of the hand-scoring of student responses for polytomous items. One of these processes was double-blind scoring of all student responses. Results of the double-blind scoring were used during scoring to identify Scorers who required retraining or other intervention, and are presented here as evidence of the reliability of the NYSAA. A summary of the interrater consistency results is presented in Table 10-4. Results in the table are collapsed across the tasks by content area and grade. The table shows the number of included scores, the percent exact agreement, the correlation between the first two sets of scores, the mean absolute difference between scores that did not have exact agreement, and the standard deviation of these absolute differences. This same information is provided at the item level in Appendix E.

Table 10-4. 2014–15 NYSAA: Summary of Interrater Consistency Statistics Collapsed Across Items by Content Area and Grade

Subject	*Grade*	*N*	*Overall Interrater Percent Exact Agreement*	*Overall Interrater Correlation Agreement*	*Overall Mean Absolute Difference for Non-Exact*	*Overall S.D. Absolute Difference*
English Language Arts	3	2,560	98.55	1.00	2.88	2.80
	4	2,971	98.35	1.00	2.19	2.58
	5	3,050	98.39	1.00	2.75	1.97
	6	2,993	98.40	1.00	2.48	2.42
	7	3,161	98.67	1.00	3.65	3.07
	8	3,140	98.85	1.00	3.34	2.50
	High School	2,850	98.53	1.00	2.57	2.65
Mathematics	3	2,546	98.90	1.00	2.35	2.40
	4	2,992	98.36	1.00	2.50	2.46
	5	3,050	98.07	1.00	3.02	2.86
	6	3,000	97.47	1.00	2.49	2.38
	7	3,186	97.77	1.00	2.17	1.80
	8	3,180	98.40	1.00	2.62	2.83
	High School	2,809	97.76	1.00	3.08	2.83
Science	4	1,189	98.65	0.99	4.88	3.80
	8	1,288	98.29	1.00	2.65	2.41
	High School	1,161	97.93	1.00	2.60	2.12
Social Studies	High School	1,153	98.87	1.00	2.43	1.73

CHAPTER 11 COMPARABILITY (SCALING AND EQUATING)

11.1 COMPARABILITY OF SCORES ACROSS YEARS

In administering the New York State Alternate Assessment (NYSAA), teachers select Extensions or Alternate Grade-Level Indicators (AGLIs), following the Test Blueprints. Use of the Extensions or AGLIs and Blueprints ensures that the assessment, as it is administered, is appropriate for the individual needs of the student being assessed and that the standards required are covered. The process enables teachers to customize the assessment for individual students while at the same time ensuring comparability across years through the use of the same Blueprints, Extensions/AGLIs, and Scoring Rubrics from year to year. Additionally, comparability is ensured through the scoring process. Teachers use the same Scoring Rubric for scoring datafolios each year, and scoring occurs at regional Scoring Institutes that all follow the same Scoring Training program and Scoring Procedures, as well as the standard scoring Quality Control Processes, as described in Chapter 8. Additional processes to ensure across-year comparability include calculation of reported scores and categorization into achievement levels, as described below.

11.1.1 Standard Setting

Standard setting was conducted in June 2014 to establish cut scores for the scaled total scores for each alternate performance level in English Language Arts (ELA) and mathematics, Grades 3 through 8 and high school; in science, Grades 4 and 8 and high school; and in high school social studies. To ensure continuity of score reporting across years, the cuts that were established at the standard-setting meeting will continue to be used in future years, until it is necessary to reset standards. The scaled total score cutpoints for the NYSAA as established via standard setting are presented in Table 11-1.

Table 11-1. 2014–15 NYSAA: Cut Scores on Reporting Scale by Content Area and Grade

Content Area	*Grade*	*Scaled Score Cuts*			*Scaled Score*	
		Cut 1	*Cut 2*	*Cut 3*	*Minimum*	*Maximum*
English Language Arts	3	426	449	491	400	525
	4	421	445	489	400	525
	5	424	444	484	400	525
	6	425	447	490	400	525
	7	429	443	486	400	525

continued

Content Area	Grade	Scaled Score Cuts			Scaled Score	
		Cut 1	Cut 2	Cut 3	Minimum	Maximum
English Language Arts	8	431	442	480	400	525
	High School	427	449	490	400	525
Mathematics	3	432	450	493	400	525
	4	428	445	488	400	525
	5	423	441	485	400	525
	6	422	445	485	400	525
	7	425	447	492	400	525
	8	426	443	483	400	525
	High School	425	446	496	400	525
Science	4	559	567	589	550	600
	8	559	569	591	550	600
	High School	559	569	591	550	600
Social Studies	High School	559	570	586	550	600

Table F-1 in Appendix F shows performance level distributions for 2014–15 by content area and grade.

11.1.2 Reported Scores (Cumulative Distributions)

Students' entry scores are calculated based on their Level of Accuracy and Level of Complexity scores for each of the final date of student performance of the Extensions or AGLIs in a given entry. The overall score is then the sum of the entry scores. Using this formula, there may be multiple ways that a student can attain a given total score.

Graphs of the cumulative reported raw score distributions for 2014–15 are provided in Appendix G. As the curves move to the right, they represent an increase in performance.

11.1.3 Performance Level Distributions

Appendix F shows the percentages of students earning scores at each performance level. A score of No Score (NS) is designated if a datafolio does not adhere to the administration guidelines. (Complete information regarding scoring can be found in the two scoring documents titled Steps for Scoring 2014–15 NYSAA Datafolios and Decision Rules for Scoring 2014–15 NYSAA Datafolios.) The percentages are presented by grade, content area, and performance level.

11.2 Linkages Across Grades

In developing the NYSAA, a content-based approach for addressing continuity across grades was implemented. Specifically, issues of continuity were addressed in the following processes: (1) development, (2) administration, and (3) standard setting.

As explained in Chapter 4, the Extensions and AGLIs describe the content to be included in students' instructional programs for each grade level. The Extensions in ELA and mathematics are

based on the Common Core Learning Standards (CCLS), and AGLIs in science and social studies are based on the core curriculum's Grade-Level Expectations, but have been reduced in depth and breadth. The Extensions and AGLIs are designed to follow a developmental continuum of skills that increases across grades. Each Assessment Task must align to the Extension or AGLI, and each is designed to measure grade-specific content and skills. These Assessment Tasks and the Extensions or AGLIs, along with the Test Blueprints, were designed to mirror the developmental continuum reflected in the Extensions and AGLIs and to ensure that each datafolio builds on the appropriate knowledge and skills, thereby reflecting the desired continuity across grades.

During administration, the Test Blueprint serves teachers as a guide to selecting Extensions and AGLIs that are appropriate for a given student. In addition, teachers must select Assessment Tasks that are aligned with the chosen Extensions and AGLIs. As with other aspects of the development and administration of the NYSAA, use of the Test Blueprints and the Extensions and AGLIs ensures that the student is being assessed at a level that is appropriate for his or her individual needs and that the Extensions or AGLIs and Assessment Tasks to which students are exposed follow a developmentally appropriate continuum from year to year. Thus, linkages across grades are built into the design of the datafolio.

Finally, the continuity of the NYSAA across grades was further verified through the standard-setting procedures. The achievement level descriptors used for standard setting were based on the student expectations as delineated in the Extensions and AGLIs. Proficiency across grades, therefore, was expected to follow the developmental continuum established by the Extensions or AGLIs, and thus to reflect a higher level of cognition as the grades increased.

CHAPTER 12 VALIDITY

12.1 PROCEDURAL VALIDITY

To ensure the consistency of the information given to teachers across New York State, sets of documents and training programs were developed and distributed statewide. New York State has a group of Alternate Assessment Training Network Specialists (AATN Specialists) and Score Site Coordinators (SSCs) who present a turnkey training provided to them by the New York State Education Department (the Department) and Measured Progress.

For the administration of the New York State Alternate Assessment (NYSAA), the materials included the following:

- *NYSAA Administration Manual*: This document contains all of the guidelines and specific requirements of the NYSAA; all of the forms required to be used in the datafolio; and the Test Blueprints, Extensions for English Language Arts (ELA) and mathematics, Alternate Grade-Level Indicators (AGLIs) for science and social studies, and Assessment Tasks for each Extension or AGLI for each grade level and content area.
- Training program DVD: The entire Administration Training program that is used with teachers is contained in this recorded program. All AATN Specialists are required to use the DVD in its entirety to train teachers. It ensures that the exact same message is imparted statewide.
- Training program slides and handouts: All slides and handouts developed by the Department and Measured Progress are required to be used by the AATN Specialists while training teachers. The handouts include slide printouts and Guided Practice activities.

For the scoring of the NYSAA, the materials included the following:

- Steps for Scoring 2014–15 NYSAA Datafolios and Decision Rules for Scoring 2014–15 NYSAA Datafolios: These are the two main documents used to guide the process for scoring each datafolio (see Appendices B and C).
- Training program DVD: The entire Scoring Training program that is used with Scorers is contained in this recorded program. All SSCs and AATN Specialists are required to use the DVD in its entirety to train Scorers. It ensures that the exact same message is imparted statewide.
- Datafolio practices and qualifiers: All Scorers must complete the three practice samples provided and then must qualify by scoring datafolio samples. All Scorers are qualified using calibrated materials that were initially identified during a Benchmarking process.

12.2 CONTENT VALIDITY

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) notes that an important part of establishing test validity is ensuring that a close, substantive relationship exists between a test's content and the underlying construct that it is intended to measure. The *Standards* further elaborate that the test content refers to the "themes, wording, and format of the items, tasks, or questions on a test. Administration and scoring may also be relevant to content-based evidence" (2014, p. 14). In addition to describing the content in detail, content validity evidence must, of course, relate the content to the construct that the test is intended to measure. One important approach in this regard, mentioned in the *Standards,* is the use of "expert judgment of the relationship between parts of the test and the construct" (2014, p. 14).

The New York State (NYS) Learning Standards provide the framework for the New York State Testing Program, including the NYSAA. For ELA and mathematics, the standards are from the NYS Common Core Learning Standards (CCLS). For science and social studies, the standards are from the NYS Learning Standards. These standards are the constructs that are intended to be measured by the NYSAA. Chapters 4 through 6 of this report describe in detail the development and design of the content for the NYSAA, with special emphasis on the relationship of the test content to the standards. Chapter 8 provides a detailed description of the scoring process for the NYSAA, again emphasizing that the procedures used ensure strong adherence to the standards. Another important component of the Scoring Procedure is the standard-setting process, in which expert judgment is used to set the scores on the test that correspond to different levels of classification of student achievement relative to the standards. The Standard Setting Report documenting the June 2014 standard-setting meeting describes the rigorous procedures that were followed to ensure that the content-related aspects of the standard setting maintained a strong, substantive alignment with the standards.

As shown from the above definition of construct validity and in the descriptions of the contents of Chapters 4, 5, 6, and 8 of this report, a complete description of the content validity of the NYSAA is available to the reader.

12.3 CONSEQUENTIAL VALIDITY

Beginning in 1997, the Department began discussions on how to provide students who have severe cognitive disabilities access to the general education standards. To that end, an advisory committee made up of New York State Special Education Teacher and general education teacher committees was formed. Its goal was to develop a handbook that would provide teachers with an alternate pathway for this group of students to gain access to the NYS Learning Standards. On July 17, 1997, the New York State Board of Regents endorsed a set of Alternate Performance Indicators (APIs)

that were linked to the NYS Learning Standards. The purpose of the APIs was to provide teachers with a way of teaching academic content to students with severe cognitive disabilities. The final manual, titled *The Learning Standards and Alternate Performance Indicators for Students with Severe Disabilities*, was published in 1998 and distributed statewide.

As mandated in the reauthorized Individuals with Disabilities Education Act of 1997 (IDEA of 1997), states were required to have an alternate assessment in place by July 2000 for those students who could not participate in the general education assessments, even with accommodations. Because of the groundbreaking work already done, the Department, in collaboration with Measured Progress and under the guidance of the advisory committee, endorsed the use of the APIs in 1997 as a way to measure the knowledge, skills, and understanding of students with severe cognitive disabilities against the NYS Learning Standards. The advisory committee concluded that all students must be given the opportunity to achieve the NYS Learning Standards, but that not all standards are appropriate for this group of students, which was in line with the intent of the IDEA of 1997. It was understood that this group of students would be assessed against APIs because of their inability to participate in the general assessments, even with accommodations. The APIs, while based on the NYS Learning Standards, were, by their very nature, functional and limited to students with severe cognitive disabilities. They reflected what was determined to be appropriate for this group of students. They were not grade specific, nor were they aligned to grade-level content. The Committees on Special Education (CSEs) determined which students were appropriate for inclusion in the NYSAA, based on several strict criteria, and decided on which APIs the students would be assessed. The first NYSAA was piloted between March 1998 and March 2000, with full implementation during the 2000–01 school year. The purpose of the NYSAA was to promote the inclusion of students with severe cognitive disabilities in the statewide assessment program. It was not for the purposes of Adequate Yearly Progress (AYP) as defined by the No Child Left Behind (NCLB) Act of 2001.

The following is the calendar of events that the Department followed to develop and implement its first alternate assessment.

Spring 1998	Conduct regional training for teachers on the APIs
March 1998–March 2000	Develop and pilot the alternate assessment system
March–June 2000	Provide information and training on the alternate assessment system
July 2000	Implement a statewide alternate assessment system as required by the IDEA of 1997
June 2001	Collect data and report student scores to the public

The Department and its Special Education Teacher and general education teacher committees were committed to building an assessment and accountability system that included students with

severe cognitive disabilities. New York State was one of the first states to engage teachers, administrators, policymakers, and others in these important discussions, and it did pioneering work in the early years of alternate assessment.

With the reauthorization of the NCLB Act of 2001, students in every state are being held to high levels of academic achievement, including students with severe cognitive disabilities. The original NYSAA tested students in Grades 4 and 8 and high school in the content areas of ELA, mathematics, science/health, and social studies. Based on new grade-level testing requirements in NCLB, in September 2005 the Department began to implement a revised NYSAA that included Grades 3 through 8 and high school in the content areas of ELA, mathematics, science, and social studies. The students were assessed against the original APIs; however, the format and the number of APIs assessed were modified. Table 12-1 outlines the revised NYSAA.

Table 12-1. 2014–15 NYSAA: Revised NYSAA
Grades 3–8 and High School

Datafolio Component	*Grade Equivalents*	
	Anchor *4, 8, and High School*	*Expanded* *3, 5, 6, and 7*
Table of contents	✓	✓
Student Page	✓	✓
One Entry Cover Sheet for each content area	English Language Arts, mathematics, social studies, science	English Language Arts, mathematics
One Data Summary Sheet for each content area	4 (one for each content area above)	2 (one for English Language Arts, one for mathematics)
Verifying evidence per API	1 piece per API in each content area	3 pieces for mandatory API in English Language Arts and mathematics
Permission to tape and photograph	If applicable	If applicable
Digital Video and Audio Clip Summary form	If applicable	If applicable

During the 2005–06 testing cycle, the Department submitted its accountability documentation for peer review to the U.S. Department of Education. The results of that review required the Department to revise its alternate assessment to ensure:

- the presence of evidence of alignment between the NYSAA alternate achievement standards and the newly adopted Grade-Level Expectations;
- that students are assessed at each required grade;
- the setting of cutpoints and the development of Alternate Performance Level Descriptors (APLDs) for each grade level and content area; and

- the technical quality of the assessment, including research-based standard setting, and the production and submission of the Standard Setting Report and Technical Report.

The new assessment system had to be in place for the 2006–07 testing cycle, culminating with standard setting in June 2007.

Beginning in July 2006, the Department, in collaboration with Measured Progress, redesigned the NYSAA. The focus and purpose of the assessment is to ensure that students with severe cognitive disabilities are being provided access to the general education curriculum (e.g., Grade-Level Expectations). However, for these students, Grade-Level Expectations need to be expanded in both breadth and depth. This resulted in development of the AGLIs, which were contained in the *NYSAA Administration Manual*: Appendix H—NYSAA Frameworks.

The Department brought together groups of Special Education Teacher and general education teacher committees, including general education content specialists and Special Education Teachers, to develop the AGLIs. The groups referred to the general education Test Blueprints to determine the academic core priorities. From there, each content group reviewed the Grade-Level Expectations for each grade level and content area. The groups determined the Essences of the Grade-Level Expectations. Lastly, the groups wrote AGLIs that were aligned to the Essences of the Grade-Level Expectations. In addition to developing the AGLIs, Special Education Teacher and general education teacher committees were also brought together to develop Sample Assessment Tasks (SATs) aligned to the AGLIs. The following year, the Stakeholder groups were brought in again to further refine what was originally developed.

The new NYSAA was first administered in late fall 2006. This abbreviated administration period culminated with regional Scoring Institutes. Standard setting was conducted in June 2007, resulting in cut scores for each grade level and content area, as well as in APLDs. The cut scores were approved by the Commissioner of Education and submitted, along with the Standard Setting Report, to the U.S. Department of Education. The 2007–08 NYSAA administration was a full administration period, and was based on the refined AGLIs and SATs. This administration, too, culminated with the regional Scoring Institutes. Standard setting was conducted on the revised AGLIs in June 2008, resulting in new cut scores and updated APLDs for each grade level and content area. The Commissioner of Education approved the updated cut scores in June 2008. The intent of the AGLIs was not changed following the 2007–08 administration; therefore, the cut scores established during the June 2008 standard setting remained consistent for each grade level and content area.

The New York State Board of Regents committed to the CCSS in January 2010 and formally adopted the CCSS for ELA and mathematics in July 2010 and incorporated New York State-specific additions, creating the New York State P-12 CCLS in January 2011. The Board of Regents announced that, for students with severe cognitive disabilities, student progress on the CCLS would be measured beginning with the 2014–15 administration of the NYSAA in ELA and mathematics.

To align the ELA and mathematics NYSAA with the CCLS, the Department began work in November 2011 to develop an alternate assessment to measure the CCLS for ELA and mathematics for students with severe cognitive disabilities. The Department brought together teacher committees made up of general education content specialists and Special Education Teachers, to review the CCLS for the content identified in the new Test Blueprint, develop Essence statements to narrow the depth and breadth of the CCLS, and draft Extensions. The new Test Blueprint for the NYSAA was based on the general education Test Blueprints to determine the academic core priorities. The Test Blueprints for the NYSAA that measure the Extensions to the CCLS in ELA and mathematics were approved in spring 2012. The draft Essences and Extensions were reviewed extensively during the summer of 2012, and draft documents were posted for public comment in September 2012. In October 2012, the committees were reconvened to review the revisions to the Essences and Extensions, and to draft Assessment Tasks to measure student performance of the CCLS. Following the meeting, the draft Assessment Tasks were reviewed and vetted by content and Special Education Teachers, and then were posted for public comment from December 2012 to January 2013. Public comments from the first review and the second review were incorporated, as appropriate, into the draft Extensions and draft Assessment Tasks.

The administration procedures for the NYSAA were revised to a new test design that emphasized a continuum of student performance and intentionally focused on increasing the validity of test scores. The procedures were streamlined, where possible, to reduce teacher clerical errors. The 2014–15 administration procedures applied to all content areas. This was done so that teachers would not have to use two completely different assessment procedures. The new NYSAA was first implemented in 2014–15. The administration culminated with regional Scoring Institutes. Standard setting was conducted in June 2014, resulting in cut scores for each grade level and content area, as well as in APLDs. The cut scores were approved by the Commissioner of Education, following standard setting.

The information in this section and throughout the Technical Report provides a framework to determine the consequential validity of the NYSAA. In order to demonstrate consequential validity, the assessment should:

- provide multiple measurement occasions,
- show that student results are improving, and
- demonstrate that revisions to the NYSAA are considered based on Stakeholder feedback.

The revised NYSAA demonstrates that students are provided multiple measurement occasions as embedded in the baseline and final data collection points. Also, Stakeholder input has been critical throughout the development and revision processes.

REFERENCES

Alternate Assessment Alignment Pilot Study: Report to the State Department of Education. Prepared by Claudia Flowers, Diane Browder, and Shawnee Wakeman, University of North Carolina at Charlotte, and Meagan Karvonen, Western Carolina University, April 2006.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). Fort Worth, TX: Holt, Rinehart and Winston.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355–368.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley and Sons, Inc.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Joint Committee on Testing Practices. Available from www.apa.org/science/programs/testing/fair-code.aspx.

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

Tindal, G. (2005). *Alignment report 2: Alignment of alternate assessments using the Webb system.* Retrieved 04/16/07 from www.ccsso.org/publications/details.cfm?PublicationID=293.

Webb, Norman L. *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Council of the Chief State School Officers. December 2002.

APPENDICES

APPENDIX A—NYSAA TEST BLUEPRINTS FOR EACH CONTENT AREA

New York State Alternate Assessment (NYSAA)
Test Blueprint for English Language Arts (ELA)

ENGLISH LANGUAGE ARTS (ELA)

Strand	Sub-Strand	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Reading Standards for Literature	Key Ideas and Details		X		X			X
	Craft and structure			X				
	Integration of Knowledge and Ideas		X			X		
	Responding to Literature	X						
Reading Standards for Informational Text	Key Ideas and Details				X			X
	Craft and Structure	X					X	
	Integration of Knowledge and Ideas			X			X	
	Key Ideas & Integration of Knowledge and Ideas					X		
Writing Standards	Text Types and Purposes	X			X			X
	Production and Distribution of Writing		X			X		
	Research to Build and Present Knowledge			X			X	
Speaking and Listening Standards	Comprehension and Collaboration	X		X	X	X		
	Presentation of Knowledge and Ideas		X				X	X
Language Standards	Conventions of Standard English	X			X			X
	Knowledge of Language			X			X	
	Vocabulary Acquisition and Use		X			X		

Domain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Operations and Algebraic Thinking (OA)	X	X	X				
Number and Operations in Base Ten (NBT)	X	X	X				
Number and Operations – Fractions (NF)	X	X	X				
Measurement and Data (MD)	X	X	X				
Geometry (G)	X	X	X	X	X	X	
Ratios and Proportional Relationships (RP)				X	X		
The Number System (NS)				X	X		
Expressions and Equations (EE)				XX	X	XX	
Functions (F)						X	
Statistics and Probability (SP)					X	X	
Quantities (NQ)							X
Creating Equations (A-CED)							X
Interpreting Functions (F-IF)							X
Expressing Geometric Properties with Equations (G-GPE)							X
Interpreting Categorical and Quantitative Data (S-ID)							X

NYSAA Test Blueprint - Science
Effective with 2014–15 Administration

Two Standards are assessed for each Grade as Marked by an X				
Standards	**Chapter 5. Key Idea**	**Grade 4**	**Grade 8**	**High School***
1 – Analysis, Inquiry, and Design (Scientific Inquiry)	2- Beyond the use of reasoning and consensus, scientific inquiry involves the testing of explanations involving the use of conventional techniques and procedures and usually requiring considerable ingenuity.	X		
	3- The observations made while testing proposed explanations, when analyzed using conventional and invented methods, provide new insights into phenomena..		X	
4- Living Environment	1- Living things are both similar to and different from each other and from nonliving things.			X
	3- Individual organisms and species change over time.	X		
4- Physical Setting/ Earth Science	2- Many of the phenomena that we observe on Earth involve interactions among components of air, water, and land.			X
	3- Matter is made up of particles whose properties determine the observable characteristics of matter and its reactivity.		X	

*Note: See the Core Curricula for Science at http://www.p12.nysed.gov/ciai/cores.html#MST.

NYSAA Test Blueprint - Social Studies (HS only)
Effective with 2014–15 Administration

Two Standards are assessed for each Grade as Marked by an X		
Standards	*Units*	**High School**
1- US History	2 - Constitutional Foundations	X
2- World History: Global History and Geography	8 - Global Connections and Interactions	X

See the Core Curricula for Social Studies at:
http://www.p12.nysed.gov/ciai/cores.html#SOCIALSTUDIES

APPENDIX B—NYSAA Scoring Procedures

Procedures for Scoring NYSAA Datafolios 2014–15

- **Follow the steps outlined below to review each NYSAA datafolio.**
- Review the documentation to determine the answer to the question/statement for each step.
- If a discrepancy is not addressed in this document, **consult your Table Leader**.
- Prior to the Scorer recording the error, a Table Leader **MUST** review and confirm all issues that may result in a **"No"** for any of the three Connections questions, a **"No Score"** for a date(s), and/or an **adjustment(s)** to the Data Summary Sheet (DSS).

1. Student Demographics, Scorer ID, Scoring Institute Code

a) Is the student demographic information consistent?

Student demographic information must be consistent between the demographic label (from the RIC), Student Page (in datafolio), and Scannable Score Document. If discrepant or if scannable is missing, **consult the Table Leader**. Record Scorer comment **A** at the bottom of the Scorer Worksheet.

b) Apply label in the upper left corner on each page of the Scorer Worksheet. If a label is not available, transcribe the information from the Student Page to the Scorer Worksheet.

c) Fill in your Scorer ID and the Scoring Institute Code.

Enter your 3-digit Scorer Identification Number and 6-digit Scoring Institute Code in the upper-right corner of the Scorer Worksheet.

d) Is the student's DOB within the range indicated on the Student Page for the grade assessed?

i. If Measured Progress ProFile™ was used - accept the grade level as correct.

ii. Note: If a **DOB** is **outside the range specified** for any grade level, **consult your Table Leader**.

If **YES** →	Mark the grade assessed in the upper right corner of the Scorer Worksheet.	
If **NO** → ***Consult the Table Leader***	**Wrong grade level** was assessed.	<u>*Record:*</u> • Extension/AGLI code **00099** • "**N**" for **No** for **all Connections questions** for each Extension/AGLI within the content areas that should have been assessed • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **1**

e) Are there any Testing Accommodations listed on page two of the Student Page?

If **YES** →	Transcribe any Testing Accommodations to the Scannable Score Document.
If **NO** →	Continue to review and score the assessment. If page two of the Student Page is missing or incomplete, continue to review and score the assessment.

f) Was a Collegial Review month indicated on the Student Page?

If **YES** →	<u>*Record:*</u> • "**Y**" for **Yes** for "Was a collegial review of this datafolio conducted?" on the Scannable Score Document.
If **NO** →	<u>*Record:*</u> • "**N**" for **No** for "Was a collegial review of this datafolio conducted?" on the Scannable Score Document. • Scorer comment **B**, at the bottom of the Scorer Worksheet Continue to review and score the assessment.

g) Set aside the Scannable Score Document until all content areas have been reviewed and scored.

Datafolios are scored in order: ELA, mathematics, science and social studies. Review the entire datafolio to determine if anything is out of order. **Do not reorganize the datafolio**. A datafolio with out-of-order documents can be scored.

2. Review Sequence of Documentation for Content Area

a) Are the required Data Summary Sheets (DSSs) present, one for each standard assessed? (refer to DSS Titles in the upper-right corner [e.g., Looking at ELA and math, do you have 1, 2, 3, 4, and 5 in titles; in science and social studies, do you have 1 and 2])

If **NO** → ***Consult the Table Leader***	DSS **missing**.	*Record:* • Extension/AGLI code **00099** • "**N**" for **No** for **each Connection question** for the missing Extension/AGLI • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **2** Proceed to the next Extension/AGLI that has a DSS or next content area if all DSSs are missing.
	Two or more DSSs are found for the **same Standard** and there are fewer than the required number of Extensions/AGLIs assessed.	• Review and score the first DSS and assessment documentation for the Extension/AGLI assessed *For the missing Extension/AGLI, record:* • Extension/AGLI code **00099** • "**N**" for **No** for **each Connection question** • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **5** Proceed to the next Extension/AGLI or the next content area.

b) Are the DSS forms in order?

Confirm that the DSSs are in the correct order using the titles (upper right corner) on each form.

If **NO** → ***Consult the Table Leader***	Documents are **out of order**.	• Consider documentation that is out of order and score the assessment **in the correct order**. **Do not reorganize the datafolio**. *Record:* • Scorer comment **C** Proceed to Step 3.

3. Is demographic information on DSS complete and accurate when compared to the Student Page?

If **NO** → ***Consult the Table Leader***	Demographic information is **discrepant or incomplete**.	• Transcribe information from the Student Page to the DSS **in red ink**. *Record:* • Scorer comment **D** Proceed to Step 4.

4. Extension/AGLI from Grade Level

Is the Extension/AGLI indicated on the DSS from the student's assessed grade?

If **YES** →	• Measured Progress ProFile™ was used to complete datafolio documentation **or** • Extension/AGLI is from grade.	*Record:* • **Extension/AGLI code** (5-digits) • **"Y"** for **Yes** for "**Extension/AGLI from Grade**" Proceed to Step 5.
If **NOT SURE** → ***Consult the Table Leader***	Extension/AGLI is **missing** on the DSS, **but can be located** on VE.	If Extension/AGLI code or text is **found on the VE**, and code/text matches the Frameworks, transcribe the information to the DSS **in red ink** and continue to review and score the assessment. *Record:* • **Extension/AGLI code** (5-digits) • **"Y"** for **Yes** for "**Extension/AGLI from Grade**" • Comment **D2** Proceed to Step 5.
If **NO** → ***Consult the Table Leader***	DSS for a standard includes an Extension/AGLI from an inappropriate grade level.	*Record:* • Extension/AGLI code **00099** • "**N**" for **No** for **each Connection question** • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **4** Proceed to next Extension/AGLI or content area.

5. Task Connects to Extension/AGLI

Does the Assessment Task (AT) documented on the DSS clearly connect to the Extension/AGLI? (reference Extension and AT codes)

If Measured Progress ProFile™ is not used and the Assessment Task is hand written, Scorer must verify code and text against the AT in the Frameworks to ensure text is exact and AT is from the same Level of Complexity as the Extension/AGLI.

If **YES** →	• Measured Progress ProFile™ used to complete DSS **or** • Assessment Task **clearly connects** to selected Extension/AGLI.	*Record:* • **"Y**" for **Yes** for "**Task Connects to Extension/AGLI**" Proceed to Step 6.
If **NOT SURE** → ***Consult the Table Leader***	Assessment Task (code/text) is **missing** on the DSS, but can be located on VE.	If the Assessment Task **code or text** is **found on the VE** and the code/text matches the Frameworks, transcribe the information to the DSS **in red ink** and continue to review and score the assessment. *Record:* • **"Y"** for **Yes** for "**Task Connects to Extension/AGLI"** • Comment **D3** Proceed to Step 6.
If **NO** → ***Consult the Table Leader***	Assessment task **does not connect to Extension/AGLI**.	*Record:* • "**N**" for **No** for "**Task Connects to Extension/AGLI**" and **remaining Connection question** • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **6a** Proceed to next Extension/AGLI or content area.
	Assessment task is **missing** and cannot be located on the VE (either evidence itself or VE label).	*Record:* • "**N**" for **No** for "**Task Connects to Extension/AGLI**" and **remaining Connection question** • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **2** Proceed to next Extension/AGLI or content area.

6. Verifying Evidence (VE) Connects to Task (confirming only if VE is connected, not whether VE is valid)

a) Is evidence for both the baseline and final data point found behind the DSS? (VE for two separate dates)

Note: A single DCS may be considered as two pieces of VE. A calendar/chart can only be submitted as one piece of VE.

If **NOT SURE** → ***Consult the Table Leader***	If VE appears to be missing	Review the entire datafolio to determine if missing piece of VE is misplaced. If VE is misplaced, leave it where it is found, review and score the assessment. Do not reorganize the datafolio. *Record:* • Scorer comment **C** Proceed to Step 6b
If **MORE THAN TWO PIECES OF VE** →	If **more than baseline** and **final** VE are included Note: Do not confuse this with a student work product that is multiple pages, or with supporting evidence.	• Only evidence for the two dates (baseline and final) documented on the DSS can be considered. • Also, if one or both pieces of evidence for this Extension/AGLI are found to be invalid (Step 8), other evidence cannot be considered in its place. • If date(s) on VE are discrepant with DSS, use VE for earliest date as baseline and latest date for final. *Record:* • Scorer comment **G** Proceed to Step 6b using only VE for baseline and final.
If **NO** → ***Consult the Table Leader***	**Only one piece of evidence is found**	*Record:* • "**N**" for **No** for "**VE connects to task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **7a, 7b** as appropriate Proceed to next Extension/AGLI or content area.

b) Does the evidence for each data point connect to the Assessment Task documented on DSS?

- Data Collection Sheets (DCS) **must include step**, **trial** or **time segment** information
- Digital video and audio evidence must be accessible and can be reviewed by Scorer
- To connect, **each** piece of evidence, on its own, must:
  - ✓ Meet the **intent of task** by demonstrating the student's skill on the assessed task
  - ✓ Not include information (e.g., directions, items) that conflicts with the vocabulary of task (e.g., main character = main character, main character ≠ important character; reference content glossary to confirm)
  - ✓ Demonstrate any **plural** in the task (if parenthesis around "s," plural requirement is optional)
  - ✓ Demonstrate any **AND** in the task by demonstrating **all** elements of the assessed task

If **YES** →	Evidence for both the **baseline** and **final data point demonstrate** the above **criteria** on their own.	Proceed to Step 6c.
If **NOT SURE** → ***Consult the Table Leader***	If connection is not clear for either the **baseline** or **final** data point	If Table Leader confirms connection is demonstrated *Record:* Scorer comment E1 or E2
If **NO** → ***Consult the Table Leader***	Evidence for baseline and/or final data point **does not connect to Assessment Task.**	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **8a, 8b, 8c,** or **8d** Proceed to next Extension/AGLI or content area.

	DCS includes a **single step or time-segment** that **does not clearly document** the student performance for the assessed task (e.g., plural or AND) and there is no other information or notation to clarify how task was conducted.	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **3a, 3b, or 3c** Proceed to next Extension/AGLI or content area. Note: it is acceptable to use a multi-step DCS to demonstrate a single step task if all requirements are met
If **NO** → ***Consult the Table Leader (cont'd)***	Assessment Task includes a **plural without parenthesis around "s."** Upon review, one or both pieces of VE **do not satisfy the plural**.	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **8d** Proceed to next Extension/AGLI or content area.
	Assessment Task includes an **AND** statement. Upon review, one or both pieces of VE **do not satisfy the AND** statement.	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **8d** Proceed to next Extension/AGLI or content area.
	DCS included as VE is **missing the step, trial** (skill/sub-skill) **or time-segment** information.	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **16d** Proceed to next Extension/AGLI or content area.
	Digital video and/or audio **malfunctioned** or the clip is **unable to be located** on the DVD and/or CD.	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **15b** or **15d** Proceed to next Extension/AGLI or content area.

c) Is the BASELINE Level of Accuracy 74% or below and calculated correctly? (below threshold)

If **YES** →	On review of the VE for **baseline**, Level of Accuracy score is correct and confirmed to be **74% or below**.	*Record:* • "**Y**" for **Yes** for "**VE Connects to Task**" Proceed to Step 7. Note: Do not record the Level of Accuracy at this step. That will be done in Step 10.
If **NO** → ***Consult the Table Leader***	Level of Accuracy score for baseline is **75% or higher** (as documented or recalculated).	*Record:* • "**N**" for **No** for "**VE Connects to Task**" • "**NS**" for **No Score** for **baseline** and **final dates** • Procedural Error comment **8f** Proceed to next Extension/AGLI or content area.
If **NOT SURE** → ***Consult the Table Leader***	Information on the VE **is missing, contradicts or does not support what is documented** for the Level of Accuracy on DSS.	• If Level of Accuracy is missing on DSS, but score can be found and verified on VE, transcribe missing percentage to DSS and continue to score the assessment. Record Scorer comment **D5.** • **If error in calculation is clear** *For example:* *Math example: "4 + 2 = 6" is marked incorrect by the teacher but is clearly correct.* **Adjust** the baseline score on the DSS in red ink. If recalculation forces baseline score to 75% or higher, proceed to **No** above. • If scorer disagrees, or correct answer can be debated but **cannot be clearly resolved** one way or the other, accept the percentage documented. Scorer comment **K3 or N1, N2,** or **N3.** Proceed to Step 7.

7. Dates of Student Performance on the DSS

Are dates recorded on the DSS for the baseline and final data points within the 2014–15 NYSAA administration period (September 29, 2014–February 27, 2015)?

If **NO** → ***Consult the Table Leader***	One or both dates of student performance is **missing from DSS, but can be determined** from VE and dates are within the administration period.	• Transcribe date(s) from VE to the DSS **in red ink**. *Record:* • Scorer comment **J**
	One or more dates of student performance **cannot be determined** from VE or one or more dates on DSS are **outside** the **administration period**.	*Record:* • "**NS**" for **No Score** for **date(s)** in question • Procedural Error comment **11a** or **11b** Review the remaining date(s); proceed to next Extension/AGLI, or content area.

8. Determining the validity of Verifying Evidence (VE) and Supporting Evidence (SE)

a) Are the THREE required elements clearly documented on each piece of VE and match the information on the DSS? (Verifying Evidence and Supporting Evidence)

Required elements may be handwritten or printed on the actual VE, on a VE label that is affixed to the VE, or a combination of both. **A student may record** his or her name and/or the date on work products. It is acceptable for only the student's **first name** to be documented on the VE.

If **YES** →	**Student name**, **date of student performance** and **Level of Accuracy** are clearly documented.	Proceed to Steps 8b, c, e, and/or f depending on type of evidence.
If **NO** → ***Consult the Table Leader***	One or more required elements on the VE and/or VE label is **discrepant** with the DSS. *Note: Use chart below*	• Adjust the required element(s) on the DSS **in red ink**. *Record:* • Scorer comment **K1, K2,** or **K3** Proceed to Steps 8b, c, e, and/or f. *Note: Do not make any marks on VE or VE labels*

The following …	**Supersedes…**
Required elements documented by the teacher **on the VE**…	The DSS and the VE label
Required elements **on the VE label**……..	The DSS
Teacher recorded information…….……	Student recorded information on VE

If **NO** → ***Consult the Table Leader***	One or more required elements is **missing** from VE (VE itself or VE label) or VE label is **not affixed** to VE.	*Record:* • "**NS**" **for No Score** for **that date** • Procedural Error comment **12a, 12b,** or **12c** Review remaining date, or proceed to next Extension/AGLI, or content area.

b) Student Work Product

- **Original, no** photocopies of student responses, correction fluid/tape or white/black out.
- Students may use assistive technology, computers, and/or interactive white board systems (e.g., SMART board) to complete the student work product.

If **YES** →	Continue to review the other piece of VE submitted or proceed to Step 9.	
If **NO** → ***Consult the Table Leader***	Work product is **not original** (i.e., photocopies of student responses, correction fluid/tape, black out; teacher erasures).	*Record:* • "**NS**" for **No Score** for **that date** • Record Procedural Error comment **13** Review remaining date, or proceed to next Extension/AGLI, or content area.

c) Data Collection Sheet

To be valid the DCS must have

- a minimum of **three dates** of documented student performance,
- **one piece of SE** for **each date transcribed** to the DSS as VE,
- **staff initials** recorded for each date on the DCS, and
- **Supporting evidence** may be an Observer Verification Form (OVF) OR another type of VE.

NOTE: It is acceptable for DCS to be considered as VE for only the baseline, only the final, or both the baseline and final data points (and another type of VE submitted for the other data point). In all cases, the DCS must include at least 3 dates of data and meet all other requirements.

If **YES** →	Continue with Step 8d below; review each submitted piece of SE individually.	
If **NO** → ***Consult the Table Leader***	**SE is missing** for **any date transcribed** to DSS as VE.	*Record:* • "**NS**" for **No Score** for **the date(s) transcribed from the DCS** to the DSS • Procedural Error comment **16c** Review remaining date or proceed to next Extension/AGLI or content area.
	Fewer than three dates are documented on the DCS.	*Record:* • "**NS**" for **No Score** for **all dates on the DCS transcribed** to the DSS • Procedural Error comment **16a** Review remaining date, or proceed to next Extension/AGLI, or content area.
	Staff initials are missing from DCS for **any date with student performance data**.	*Record:* • "**NS**" for **No Score** for **all dates on the DCS transcribed** to the DSS • Procedural Error comment **16b** Review remaining date or proceed to next Extension/AGLI or content area.

d) Supporting Evidence

1. **Student Work Product -** Review Steps 8a and b to determine if student work product is valid SE.
2. **Photographs -** Review Steps 8a and e to determine if photographs are valid SE.
3. **Digital video and/or audio clip -** Review Steps 8a and f to determine if digital video and/or audio clip is valid SE.

If **YES** →	Continue to review the other piece of SE submitted or proceed to Step 9.	
If **NO** → ***Consult the Table Leader***	Student work product, photographs, or digital video or audio clip is **invalid per Step criteria**.	*Record:* • "**NS**" for **No Score** for **that date** • Appropriate Procedural Error comment indicated in Steps 8a, b, e, or f Review the other piece of SE submitted for remaining date or proceed to next Extension/AGLI or content area.

4. **Observer Verification Form (OVF)**
   - Review Step 8a and OVF criteria below to determine if OVF is valid SE.
   - NOTE: Only a DCS requires SE. Ignore an OVF submitted in support of original student work, photographic, digital video, or audio evidence.

Criteria for an OVF
An OVF is <u>invalid</u> if: • student name, date of student performance, and/or Level of Accuracy are missing; • supplementary school personnel signed as the observer (e.g., teacher aide or teacher assistant)*;* • the person collecting the data (initials on DCS) also signed the OVF as the observer for that date *(*confirmed by comparing initials and staff key information*)*; • more than one date of student performance is documented on a single OVF; • the observer's signature and/or title is not included, cannot be confirmed; • the observer's signature date is missing, or is not the same date the task was observed; or • date of performance or Level of Accuracy on OVF is discrepant with DCS

If **NO** → *Consult the Table Leader*	OVF is **invalid per one or more criteria** listed in the bullets above.	*<u>Record:</u>* • "**NS**" for **No Score** for **that date** • Procedural Error comment **12a–c** or **17a–e** Review the other piece of SE submitted for remaining date or proceed to next Extension/AGLI or content area.
	Observer's **title is missing** from OVF, **but can be confirmed** from another OVF in the datafolio.	Score the assessment. *<u>Record:</u>* • Record Scorer comment **M** Continue to review the other piece of SE submitted or proceed to Step 9.
	Observer's **title is missing** from OVF and **<u>cannot</u> be confirmed** from another OVF in the datafolio.	*<u>Record:</u>* • "**NS**" for **No Score** for **that date** • Procedural Error comment **17b** Review the other piece of SE submitted for remaining date or proceed to next Extension/AGLI or content area.

e) Photographs

Photographic evidence must be

- a minimum sequence of **three photographs** of the student performing the task,
- a minimum of **one caption** describing the sequence, and
- the sequence must occur on a **single date**.

If **YES** →	Continue to review the other piece of VE submitted or proceed to Step 9.	
If **NO** → ***Consult the Table Leader***	**Fewer than three photographs** are submitted of the student performing the task.	*<u>Record:</u>* • "**NS**" for **No Score** for **that date** • Procedural Error comment **14d** Review remaining date, or proceed to next Extension/AGLI or content area.
	No caption is found.	*<u>Record:</u>* • "**NS**" for **No Score** for **that date** • Procedural Error comment **14c** Review remaining date or proceed to next Extension/AGLI or content area.
	No date or multiple dates are found on the evidence.	*<u>Record:</u>* • "**NS**" for **No Score** for **the date** • Procedural Error comment **14a** or **14b** Review remaining date or proceed to next Extension/AGLI or content area.

f) Digital Video/Audio Clip

Video/Audio Clip must:

- be **90 seconds** or less (excluding markers) and
- contain at least one **recorded marker** with the Student's **name**, **date** of student performance, and **Level of Accuracy**.

If **YES** →	Continue to review the other piece of VE submitted or proceed to Step 9.	
If **NO** → ***Consult the Table Leader***	Clip duration is **longer than 90 seconds** and it is *unreasonable* to review entire clip.	*Record:* • "**NS**" for **No Score** for **that date** • Procedural Error comment **15c** Review remaining date, or proceed to next Extension/AGLI, or content area.
	Student's name, date of student performance, and Level of Accuracy, **are not recorded** on the clip in any manner. Note: VE label on DVD/CD case or box is not acceptable.	*Record:* • "**NS**" for **No Score** for **that date** • Procedural Error comment **15a** Review remaining date or proceed to next Extension/AGLI or content area.

9. Were any supports provided that guided the student to the correct answer?

Scorers must review each piece of VE to consider whether any documentation guided the student to the correct answer. Refer to the chart below for details.

Actions That Result in an Administrative Error

Templates or other formats are provided that give or lead the student to the answer. For example:

- the verifying evidence is a sequencing worksheet that contains three boxes that state "First," "Next," "Last"; the student response choices are pictures that contain the words "First," "Next," "Last."
- the verifying evidence is a number line on which the student must provide missing numbers, but the correct number is provided as a shaded or dotted number in the spot and the student has to put a sticker of the number on the spot.

If **YES** → ***Consult the Table Leader***	If **VE for baseline** and/or **final data point** includes documentation that **led the student to the answer.**	*Record:* • "**NS**" for **No Score** for **that date** • Procedural Error comment **18a** or **18b** Review remaining date or proceed to next Extension/AGLI or content area.
If **NO** →	Continue to review and score the assessment.	

10. Is the Level of Accuracy documented on DSS for the final data point calculated correctly based on VE? (Baseline Level of Accuracy was checked at Step 6c)

<table>
<tr><td>If YES →</td><td colspan="2"><u>Record:</u>
• Percentage for the Level of Accuracy for both baseline and final dates
• Yes or No for “was student prompted?” as documented on DSS for both baseline and final date. Note: do not verify prompts. Always record Yes/No for prompt, even if N/NS.</td></tr>
<tr><td rowspan="3">If NOT SURE → Consult the Table Leader</td><td>Information on the VE contradicts or does not support what is documented for the Level of Accuracy and the Scorer cannot clearly see how to correct calculation.</td><td>• If scorer disagrees, or correct answer can be debated but cannot be clearly resolved one way or the other, accept the percentage documented.
Accept the percentages the teacher documented.
• If Level of Accuracy recorded as a fraction (e.g., 1/4 instead of 25%) accept and score
<u>Record:</u>
• Percentage for final documented by the teacher
• Percentage for baseline
• Yes/No for the question “was the student prompted?” as documented on DSS for both baseline and final
• Scorer comment K3 or N1, N2, or N3
Proceed to Step 11</td></tr>
<tr><td>Level of Accuracy is missing from the DSS, but is present for a date that has valid VE.</td><td>• Transcribe percentage calculation from the VE to the DSS in red ink.
<u>Record:</u>
• Percentage from the VE for final
• Percentage for baseline
• Yes/No for the question “was the student prompted?” as documented on DSS
• Scorer comment D5</td></tr>
<tr><td>Level of Accuracy on the VE is discrepant with what is documented on the DSS.</td><td>• Adjust the percentage calculation on the DSS in red ink to match the VE.
<u>Record:</u>
• Adjusted percentage for final
• Percentage for baseline
• Yes/No for the question “was the student prompted?” as documented on DSS
• Scorer comment K3
Note: Never make changes to VE or VE labels.</td></tr>
<tr><td></td><td>Level of Accuracy was incorrectly calculated and the Scorer can clearly see how the percentage calculated can be adjusted.
Note: if Scorer cannot clearly see how to correct calculation, follow “If NOT SURE→” directions.</td><td>• If error in calculation is clear
For example:
Math example: “4 + 2 = 6” is marked incorrect by the teacher, but is clearly correct.
• Adjust the percentage calculation on DSS in red ink.
<u>Record:</u>
• Adjusted percentage for final
• Percentage for baseline
• Yes/No for the question “was the student prompted?” as documented on DSS
• Scorer comment O</td></tr>
</table>

11. Score the next Extension/AGLI

Follow Steps 3–10 for the next Extension/AGLI from the same content area.

12. Score Mathematics, Science, and Social Studies

Follow Steps 2–11 and score the remaining content areas in order for the grade assessed: mathematics (Grades 3–8 & HS), science (Grades 4, 8, & HS), and social studies (HS only).

13. Confirm your Scorer Worksheet is complete and accurate for each Extension/AGLI within each content area, including Procedural Error Comments, if applicable, and Scorer Comments

- **Extension/AGLI Code, Three Connections Questions**—Double check that a five digit Extension/AGLI code has been recorded or 00099 (if applicable); that the three Connections questions are bubbled in as "Y" or "N"; and percentages for Levels of Accuracy are recorded for both baseline and final data points.
    - Confirm Extension/AGLIs have been recorded correctly—each code is documented in the correct space.
    - Confirm Baseline Level of Accuracy is 74% or below or is an "NS" for No Score, if applicable
- **Procedural Error Comments (1–20)** — Double check that a Procedural Error Comment has been recorded on the Scorer Worksheet for each No or No Score.
- **Scorer Comments (A-O)/Positive Feedback Comments (P-W)** — Select comments from the back of the Scorer Worksheet that will clarify if something was adjusted in the datafolio and/or if something was questioned during scoring. Scorers are encouraged to also provide positive feedback to teachers.
- **No blank spaces unless the content was not assessed**.

14. Complete the Scannable Score Document

Transcribe the following data:	
From the: **Scorer Worksheet**	• **Extension/AGLI code** — 5 digits or 00099 (if applicable) • **Three Connections questions** — "Y" for Yes or "N" for No o Extension/AGLI from grade level o Task connects to Extension/AGLI o VE connects to task • **Percentages or "NS", if applicable** — Level of Accuracy for baseline and final • **Yes/No** for "was student prompted?" — even if N or NS for Extension/AGLI
From the: "**Not Tested**" **form**, if applicable	• Absent • Administrative Error • Not Enrolled • Took Another Assessment • Medically Excused
Confirm you have completed:	
From the: • **Student Page**	• Was a **Collegial Review** of this datafolio conducted? "Y" for Yes or "N" for No
	• **Transcribe the Testing Accommodations** documented on page 2 of the Student Page **to the Scannable Score Document** in the space provided.
• Complete the Scannable Score Document for each applicable content area and for any other information as directed by the SSC.	

CAUTION

Errors in transcribing Connection to Grade Level Content and Performance percentages from the Scorer Worksheet to the Scannable Score Document will directly impact the student from receiving a reportable score.
PLEASE DOUBLE CHECK ALL TRANSCRIPTIONS TO THE SCANNABLE SCORE DOCUMENT!

Other Scoring Concerns or Questions Not Addressed Directly in the Scoring Procedures

CONSULT TABLE LEADER

This table outlines other issues that may come up when scoring a datafolio. These may result in a No Score and/or adjustment to the datafolio. If any of these issues are found, consult the Table Leader for direction.

The following **may or may not result in a No or No Score.**

- **Incorrect** or **teacher-created NYSAA forms were used** (e.g., Data Summary Sheet (DSS) for the wrong grade, Student Page, or Data Collection Sheet not from NYSAA Administration Manual).
- **Student Page** is missing
- **Photocopies** (either in part or whole) or **correction fluid/tape or black out** is found on assessment documents.
- **"By" statement from Assessment Task** not demonstrated on VE.
- **Task assessed on the baseline and final data point is different**.
- Evidence is found that a **mistake in data collection was erased** on the DSS, VE, or supporting evidence and was not crossed out and initialed by the teacher.
- VE for ELA is submitted in a **language other than English**.
- Evidence is found that a **mistake in documentation by the teacher was erased** on the DSS, VE, or supporting evidence and was not crossed out, corrected, and initialed.
- **VE or supporting evidence** clearly appears to be **homework**.
- A multi-step DCS includes only a single step or a single time segment is documented on DCS.
- Assessment Task (AT) code on VE/VE label does not match Assessment Task code on DCS.

The following **may occur in a datafolio and are acceptable, providing they meet requirements.**

- **Presentation or number of items** is different between the baseline and final data point.
- Chart or calendar is submitted for a **date other than the last date recorded on the chart or calendar**.
- Verifying evidence includes items, questions, **steps not relevant** to the assessed task.
- **Extra VE** or **supporting evidence was submitted** beyond the requirements for a specific Extension/AGLI.
- **Extra DSSs** are found.
- **Dates or information printed in the header** and/or **footer** of documents completed with Measured Progress ProFile™ contradicts information recorded on the evidence or VE label.
- The **year** portion of the date in the documentation is discrepant or missing (DSS, VE, supporting evidence).

APPENDIX C—2014-15 SCORING DECISION RULES

Decision Rules for Scoring NYSAA Datafolios

(For Table Leaders)

2014–15

Rule #	Scoring Concern/Question	Decision Rule/Rationale	May appear in Step(s)
1	**Incorrect** or **teacher-created NYSAA forms were used** (e.g., Data Summary Sheet (DSS) for the wrong grade, Student Page from 2013-14, or Data Collection Sheet (DCS) not from NYSAA Administration Manual).	Incorrect Forms • If an incorrect Student Page or DSS is used but all assessment requirements can be confirmed, **score the assessment** following the Scoring Procedures. • If an incorrect DSS is used and assessment requirements cannot be confirmed, record Extension/Alternate Grade Level Indicator (AGLI) code(s) **00099,** "**N**" for **No** for **each Connection question** and "**NS**" for **No Score** for **baseline and final** of the Extension(s)/AGLI(s). Record Procedural Error comment **19**. Continue to next Extension/AGLI or content area. Teacher-created Forms • Teacher created his/her own 2014–15 forms, such as a DCS or Verifying Evidence (VE) label. If all requirements are clearly documented, **score the assessment** following the Scoring Procedures.	1-8
2	**Student Page** is **missing**	If the student demographic information (student name, date of birth) on the DSS can be used to confirm that the correct student was assessed, continue to review and score the datafolio. Direct the Scorer to record comment **A**.	1

3	**Photocopies** (either in part or whole) or **correction fluid/tape** or **black out** is found on assessment documents.	• Correction fluid/tape or black out found on page numbers, Student Page, or table of contents <u>does not directly impact</u> scores. Score the assessment following the Scoring Procedures.
		• Photocopies of the DSS, VE, or supporting evidence (either in part or in whole) or correction fluid/tape or black out found on information will directly impact student scores.
		  ○ If DSS, record "**NS"** for **No Score** for baseline and final dates.
		  ○ If VE, record "**NS"** for **No Score** for that date.
		  ○ Record Procedural Error Comment **13**
		Note: Digital photo prints in color or in black and white, computer/tablet device printouts, and interactive white board (e.g., SMART board) printouts are acceptable, since they are not photocopies.
			1–8

Assessment Task (AT)			
4	Assessment Task includes a **"by" statement** that is not demonstrated in the VE.	If any part of the Assessment Task, including a "by" statement, is not demonstrated in the Verifying Evidence and there is no teacher notation to clarify, record **"N"** for **No** to "**VE connects to Task**" and **"NS"** for **No Score** for baseline and final of the Extension/AGLI. Record Procedural Error comment **8e**. Continue to the next Extension or AGLI (e.g., grade 4 science AT42211B, "the student will distinguish between a plant and an animal by sorting a group of pictures into categories.) The *exception to this rule* is when the teacher has indicated a method of response and the student demonstrates a different method of response. This type of "by" statement is not related to the Assessment Task and a different method of response would be acceptable.	6
5	The **task assessed** on the baseline and final data points **is different.**	The same task must be assessed on the baseline and final. If the task assessed on the baseline administration is different from the final administration, record **"N"** for **No** to "**VE connects to Task**" and **"NS"** for **No Score** for baseline and final dates. Record Procedural Error comment **10**.	5–6
Verifying Evidence (VE)			
6	**Presentation** or **number of items differs** between the baseline and final VE.	Best practice for administration was to provide similar format and presentation across the baseline and final administration. For scoring 2014-15, as long as the connection of "VE to task" is clearly demonstrated, any change in format, presentation, number of items (increase or decrease), etc., should be accepted. Continue to review and score the assessment. Direct the Scorer to record comment **F**.	6
7	VE for ELA is submitted in a **language other than English**.	Record "**NS**" for **No Score** for **that date**. Record Procedural Error comment **21**. Continue to score next date.	8b–f

8	Evidence is found that a **mistake in documentation by the teacher was erased** on the DSS, VE, or supporting evidence and was not crossed out, corrected, and initialed.	• A **student may self-correct** on a student work product, which does not require a notation by the teacher. • If a teacher-made mistake is crossed out and corrected but not initialed, **score the assessment** following the Scoring Procedures. • If the mistake was crossed out but not corrected or not initialed, record "**NS**" for **No Score** for **that date**. Record Procedural Error comment **13**. • If a teacher-made erasure is confirmed, record "**NS**" for **No Score** for **that date**. Record Procedural Error comment **13**. Continue to **review and score other date** for the Extension/AGLI following the Scoring Procedures. **Note:** Documentation made by the teacher does not have to be in permanent ink.	1–10
9	**VE or supporting evidence** clearly appears to be **homework**.	• If the Student Page indicates special education programs and services at home, in a hospital, or other facility, accept what is documented by the teacher and **score the assessment** following the Scoring Procedures. • If the Student Page does not indicate special education programs and services at home, in a hospital, or other facility, record "**NS**" for **No Score** for **that date**. Record Procedural Error comment **20**. Continue to score next date.	8a–f
10	Chart or calendar is submitted for a **date other than the last date recorded on the chart or calendar**.	A chart or calendar can be submitted for only a single date. If the date on the calendar or chart is within the administration period for the 2014-15 NYSAA, accept the calendar or chart as evidence for that date. **Score the assessment** following the Scoring Procedures.	7 or 8b or 8d

11	Verifying evidence includes **items, questions, steps not relevant** to the assessment task.	• Required elements are present, all requirements for the type of VE are met, and there is no obvious error in documentation. Accept what is documented by the teacher, do not recalculate the Level of Accuracy, and **score the assessment** following the scoring procedures.
		• If all requirements are NOT met:
		  ○ Student work product: record **"NS"** for **No Score** for that date. Record Procedural Error comment **12**.
		  ○ Data Collection Sheet: record **"NS"** for **No Score** for that date. Record Procedural Error comment **12, 16,** or **17**.
		  ○ Photographic, digital video, or audio: record **"NS"** for **No Score** for **that date**. Record Procedural Error comment **12, 14,** or **15**. Continue to score the next date.
			8b-f or 10

12	A multi-step DCS includes a **single-step** task, or **single time segment** is documented on a DCS.	• All of the requirements for VE are met, the additional requirements for a DCS are met, and there is no obvious error in documentation. **Score as documented on the DCS** following the Scoring Procedures. Direct the Scorer to record comment **L**. • If a single-step task is documented on a multi-step DCS, **score the assessment** following the Scoring Procedures. • If a single time segment is documented on a DCS, **score the assessment** following the Scoring Procedures.	
13	Entry includes **extra Data Summary Sheet**(s) (i.e., separate DSS for each piece of VE).	If the same task assessed on the baseline and final dates is the same, adjust the most complete DSS with date, Level of Accuracy, and whether prompted from the other DSS (confirmed by VE) and continue to review and score the content area. Direct the Scorer to record comment **D**.	2
14	**Assessment Task (AT) code** on VE/VE label **does not match** the Assessment Task code on the DSS.	• The Assessment Task code is not a required element, if the VE connects to the task as documented on the DSS, **score the assessment** following the Scoring Procedures. • If the VE connects to the task documented on the VE, **but not** the task documented on the DSS, record **"N"** for **No** to "**VE connects to Task**" and **"NS"** for **No Score** for both the baseline and final dates. Record Procedural Error comment **8c**.	6
Dates			
15	**Dates or information printed in header** and/or **footer** of documents completed with Measured Progress ProFile™ or other web-based program contradict information recorded on the evidence or VE label.	Information printed in the header and/or footer of a document completed using the Measured Progress ProFile™ software or other web-based program (e.g., News-2-You©) cannot be considered when reviewing documentation of student performance data. **Score the assessment** following the Scoring Procedures.	1–10
16	The **year** portion of the date in the documentation is discrepant or missing (DSS, VE, SE).	When the year in a date is discrepant (e.g., Oct. 15, 2015, or 2/4/14) or missing (e.g., Oct. 15, or 2/4), but the month and day are within the acceptable parameters for the current assessment, it is considered a clerical error. **Score the assessment** following the Scoring Procedures.	7-8

APPENDIX D—SUBGROUP RELIABILITY

Table D-1. 2014–15 NYSAA: Subgroup Reliabilities—
English Language Arts

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
3	All Students	2,745	400	525	467.50	24.61	0.87	8.79
	Male	1,902	400	525	468.71	24.37	0.87	8.84
	Female	843	400	525	464.76	24.94	0.88	8.65
	American Indian/Alaskan Native	23	400	525	469.87	33.33	0.93	8.60
	Black	682	400	525	470.17	23.99	0.87	8.70
	Asian	190	400	525	468.67	27.56	0.89	8.95
	Hispanic	826	400	525	470.72	25.28	0.88	8.77
	White	972	400	525	462.85	23.03	0.85	8.82
	Native Hawaiian/Other Pacific Islander	11	400	525	475.00	25.03	0.92	7.22
	Multi	41	400	525	459.59	19.81	0.81	8.74
4	All Students	2,967	400	525	466.36	21.61	0.83	8.85
	Male	2,040	400	525	466.86	21.50	0.83	8.79
	Female	927	400	525	465.27	21.80	0.83	8.97
	American Indian/Alaskan Native	23	400	525	467.96	22.85	0.84	9.20
	Black	733	400	525	468.63	20.84	0.82	8.73
	Asian	175	400	525	465.66	20.62	0.82	8.65
	Hispanic	803	400	525	469.16	22.30	0.85	8.54
	White	1,173	400	525	463.24	21.46	0.82	9.07
	Native Hawaiian/Other Pacific Islander	19	400	525	458.00	18.14	0.81	7.89
	Multi	41	400	525	466.32	18.45	0.59	11.85
5	All Students	3,127	400	525	465.62	20.03	0.82	8.40
	Male	2,153	400	525	466.08	19.90	0.82	8.46
	Female	974	400	525	464.61	20.29	0.83	8.28
	American Indian/Alaskan Native	29	400	525	455.72	19.40	0.80	8.77
	Black	783	400	525	468.84	17.81	0.79	8.25
	Asian	157	400	525	461.55	22.20	0.85	8.63
	Hispanic	906	400	525	467.92	20.05	0.83	8.32
	White	1,211	400	525	462.51	20.54	0.83	8.55
	Native Hawaiian/Other Pacific Islander	12	400	525	469.67	19.36	0.85	7.54
	Multi	29	400	525	467.00	18.73	0.82	7.88
6	All Students	3,212	400	525	468.10	22.30	0.86	8.23
	Male	2,169	400	525	467.92	22.23	0.87	8.13
	Female	1,043	400	525	468.47	22.46	0.86	8.44
	American Indian/Alaskan Native	36	400	525	472.81	23.34	0.86	8.83
	Black	791	400	525	471.37	21.03	0.85	8.25
	Asian	172	400	525	468.09	22.12	0.88	7.74
	Hispanic	870	400	525	470.23	22.88	0.88	7.91
	White	1,301	400	525	464.47	22.30	0.86	8.45
	Native Hawaiian/Other Pacific Islander	9	400	525				

continued

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
6	Multi	33	400	525	471.45	19.00	0.72	10.04
7	All Students	3,361	400	525	463.28	24.72	0.84	9.87
	Male	2,343	400	525	463.58	24.57	0.84	9.86
	Female	1,018	400	525	462.60	25.04	0.84	9.90
	American Indian/Alaskan Native	28	400	525	462.93	29.23	0.90	9.11
	Black	816	400	525	467.83	25.50	0.85	10.00
	Asian	164	400	525	456.95	24.81	0.87	8.86
	Hispanic	865	400	525	467.12	24.34	0.82	10.25
	White	1,438	400	525	459.04	23.54	0.83	9.73
	Native Hawaiian/Other Pacific Islander	13	400	525	468.54	21.34	0.81	9.25
	Multi	37	400	525	464.76	25.28	0.85	9.88
8	All Students	3,398	400	525	460.14	24.97	0.89	8.39
	Male	2,281	400	525	459.85	24.94	0.89	8.37
	Female	1,117	400	525	460.74	25.02	0.89	8.43
	American Indian/Alaskan Native	24	400	525	463.17	16.32	0.69	9.09
	Black	867	400	525	462.57	25.33	0.89	8.58
	Asian	177	400	525	457.86	25.68	0.90	8.08
	Hispanic	910	400	525	464.70	26.70	0.91	8.20
	White	1,382	400	525	455.90	22.92	0.86	8.42
	Native Hawaiian/Other Pacific Islander	11	400	525	456.18	9.79	0.66	5.74
	Multi	27	400	525	459.00	23.79	0.89	7.98
High School	All Students	2,900	400	525	467.22	26.20	0.91	8.05
	Male	1,885	400	525	467.70	26.15	0.90	8.09
	Female	1,015	400	525	466.31	26.29	0.91	7.98
	American Indian/Alaskan Native	27	400	525	474.04	31.15	0.92	8.94
	Black	735	400	525	469.37	26.06	0.90	8.34
	Asian	141	400	525	468.16	30.65	0.93	7.91
	Hispanic	662	400	525	473.32	27.86	0.91	8.15
	White	1,302	400	525	462.64	23.89	0.89	7.84
	Native Hawaiian/Other Pacific Islander	10	400	525	484.40	31.61	0.94	7.41
	Multi	23	400	525	460.13	20.96	0.85	8.01

Table D-2. 2014–15 NYSAA: Subgroup Reliabilities—Mathematics

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
3	All Students	2,746	400	525	467.41	24.70	0.88	8.45
	Male	1,902	400	525	468.63	24.55	0.88	8.51
	Female	844	400	525	464.66	24.85	0.89	8.31
	American Indian/Alaskan Native	23	400	525	462.04	28.73	0.91	8.62

continued

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
3	Black	684	400	525	470.28	23.95	0.87	8.47
	Asian	192	400	525	466.72	27.17	0.90	8.58
	Hispanic	825	400	525	471.19	24.92	0.88	8.53
	White	970	400	525	462.65	23.69	0.88	8.30
	Native Hawaiian/Other Pacific Islander	11	400	525	474.64	26.95	0.92	7.71
	Multi	41	400	525	460.41	21.64	0.87	7.76
4	All Students	2,969	400	525	464.60	26.27	0.87	9.52
	Male	2,039	400	525	465.62	26.54	0.87	9.58
	Female	930	400	525	462.35	25.56	0.86	9.41
	American Indian/Alaskan Native	23	400	525	466.00	25.79	0.86	9.58
	Black	733	400	525	469.38	25.68	0.86	9.66
	Asian	175	400	525	465.82	25.54	0.86	9.43
	Hispanic	806	400	525	468.45	27.66	0.89	9.35
	White	1,173	400	525	458.96	24.91	0.85	9.54
	Native Hawaiian/Other Pacific Islander	18	400	525	461.56	25.58	0.85	9.86
	Multi	41	400	525	459.76	20.26	0.76	9.93
5	All Students	3,129	400	525	461.82	23.45	0.86	8.67
	Male	2,152	400	525	462.70	23.33	0.86	8.74
	Female	977	400	525	459.88	23.61	0.87	8.49
	American Indian/Alaskan Native	29	400	525	454.03	23.76	0.88	8.39
	Black	782	400	525	464.03	21.78	0.84	8.60
	Asian	156	400	525	459.28	24.18	0.88	8.39
	Hispanic	907	400	525	465.83	24.32	0.87	8.81
	White	1,213	400	525	457.87	23.08	0.86	8.63
	Native Hawaiian/Other Pacific Islander	12	400	525	464.42	23.06	0.85	8.97
	Multi	30	400	525	462.43	23.34	0.88	8.22
6	All Students	3,213	400	525	464.37	25.50	0.90	7.86
	Male	2,169	400	525	464.57	25.71	0.91	7.74
	Female	1,044	400	525	463.94	25.06	0.89	8.13
	American Indian/Alaskan Native	36	400	525	471.47	26.06	0.90	8.05
	Black	792	400	525	468.03	25.03	0.90	7.92
	Asian	172	400	525	462.46	25.65	0.90	8.27
	Hispanic	870	400	525	467.31	26.24	0.90	8.13
	White	1,301	400	525	460.11	24.70	0.91	7.57
	Native Hawaiian/Other Pacific Islander	9	400	525				
	Multi	33	400	525	468.85	22.65	0.84	9.00
7	All Students	3,364	400	525	466.82	25.47	0.84	10.19
	Male	2,344	400	525	467.13	25.31	0.84	10.14
	Female	1,020	400	525	466.12	25.85	0.84	10.28
	American Indian/Alaskan Native	28	400	525	468.29	27.09	0.88	9.43
	Black	818	400	525	471.16	26.14	0.85	10.15
	Asian	164	400	525	463.18	25.87	0.85	9.97

continued

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
7	Hispanic	864	400	525	470.10	25.23	0.83	10.30
	White	1,441	400	525	462.77	24.46	0.82	10.24
	Native Hawaiian/Other Pacific Islander	13	400	525	472.23	24.97	0.88	8.82
	Multi	36	400	525	465.19	26.94	0.86	10.10
8	All Students	3,392	400	525	457.68	25.00	0.89	8.46
	Male	2,280	400	525	457.42	25.04	0.89	8.43
	Female	1,112	400	525	458.22	24.93	0.88	8.53
	American Indian/Alaskan Native	24	400	525	459.58	17.22	0.72	9.19
	Black	865	400	525	459.63	25.26	0.88	8.76
	Asian	177	400	525	454.83	27.38	0.92	7.97
	Hispanic	909	400	525	462.57	26.53	0.89	8.72
	White	1,379	400	525	453.53	22.94	0.87	8.15
	Native Hawaiian/Other Pacific Islander	11	400	525	457.55	12.40	0.62	7.66
	Multi	27	400	525	459.96	23.38	0.89	7.74
High School	All Students	2,903	400	525	470.14	28.28	0.87	10.28
	Male	1,885	400	525	470.98	28.18	0.86	10.40
	Female	1,018	400	525	468.57	28.40	0.87	10.04
	American Indian/Alaskan Native	27	400	525	473.00	32.40	0.88	11.44
	Black	735	400	525	471.44	28.70	0.87	10.19
	Asian	141	400	525	472.01	30.97	0.89	10.07
	Hispanic	660	400	525	475.63	29.90	0.89	9.95
	White	1,307	400	525	466.40	26.24	0.84	10.46
	Native Hawaiian/Other Pacific Islander	10	400	525	485.50	30.28	0.93	8.23
	Multi	23	400	525	461.74	25.16	0.82	10.67

Table D-3. 2014–15 NYSAA: Subgroup Reliabilities—Science

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
4	All Students	2,953	550	600	576.50	10.54	0.74	5.38
	Male	2,030	550	600	576.83	10.43	0.73	5.45
	Female	923	550	600	575.76	10.76	0.76	5.24
	American Indian/Alaskan Native	23	550	600	575.61	11.22	0.86	4.18
	Black	729	550	600	578.01	10.34	0.71	5.56
	Asian	174	550	600	576.50	10.80	0.80	4.79
	Hispanic	803	550	600	577.19	10.81	0.77	5.20
	White	1,165	550	600	575.07	10.32	0.72	5.46
	Native Hawaiian/Other Pacific Islander	18	550	600	572.89	9.99	0.73	5.16
	Multi	41	550	600	578.88	8.59	0.47	6.23
8	All Students	3,382	550	600	578.83	10.48	0.70	5.75
	Male	2,273	550	600	578.67	10.64	0.71	5.72

continued

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
8	Female	1,109	550	600	579.14	10.15	0.67	5.82
	American Indian/Alaskan Native	24	550	600	580.33	11.58	0.77	5.58
	Black	864	550	600	579.34	10.19	0.71	5.52
	Asian	177	550	600	578.50	10.72	0.72	5.72
	Hispanic	902	550	600	580.36	10.70	0.68	6.05
	White	1,377	550	600	577.55	10.36	0.70	5.71
	Native Hawaiian/Other Pacific Islander	11	550	600	575.73	3.23	-1.79	5.39
	Multi	27	550	600	577.96	10.69	0.61	6.70
High School	All Students	2,896	550	600	580.43	11.59	0.77	5.59
	Male	1,880	550	600	580.54	11.47	0.77	5.54
	Female	1,016	550	600	580.23	11.81	0.77	5.68
	American Indian/Alaskan Native	27	550	600	581.37	12.80	0.83	5.20
	Black	731	550	600	581.26	11.20	0.75	5.65
	Asian	141	550	600	580.08	13.41	0.82	5.65
	Hispanic	659	550	600	582.14	11.56	0.76	5.69
	White	1,305	550	600	579.08	11.46	0.77	5.51
	Native Hawaiian/Other Pacific Islander	10	550	600	586.20	13.00	0.93	3.54
	Multi	23	550	600	580.91	9.87	0.63	6.02

Table D-4. 2014–15 NYSAA: Subgroup Reliabilities—Social Studies

Grade	Group	Number of Students	Total Scaled Score				Alpha	SEM
			Minimum	Maximum	Mean	Standard Deviation		
High School	All Students	2,886	550	600	580.17	10.99	0.80	4.89
	Male	1,875	550	600	580.42	10.83	0.80	4.90
	Female	1,011	550	600	579.71	11.26	0.81	4.87
	American Indian/Alaskan Native	26	550	600	583.38	11.34	0.81	4.99
	Black	731	550	600	581.27	10.87	0.80	4.90
	Asian	141	550	600	580.17	12.12	0.82	5.08
	Hispanic	657	550	600	582.10	11.42	0.83	4.77
	White	1,298	550	600	578.52	10.46	0.78	4.95
	Native Hawaiian/Other Pacific Islander	10	550	600	585.30	12.08	0.86	4.56
	Multi	23	550	600	577.39	9.38	0.80	4.25

APPENDIX E—INTERRATER CONSISTENCY

Table E-1. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 3

Standard	LOC	Number of Exact Matches	Number of Responses Scored Twice	Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
314	1	60	0	100.00	1.00	0.00	
	2	397	3	99.25	1.00	2.17	
	3	46	0	100.00	1.00	0.00	
322	1	367	7	98.13	0.97	4.29	
	2	113	3	97.41	0.96	2.90	
	3	20	2	90.91	0.74	2.90	
331	1	269	4	98.53	1.00	0.58	
	2	160	1	99.38	0.98	4.00	
	3	85	1	98.84	0.99	2.50	
341	1	111	1	99.11	0.98	6.70	
	2	299	3	99.01	0.94	8.33	
	3	92	0	100.00	1.00	0.00	
351	1	248	7	97.25	1.00	0.69	
	2	190	3	98.45	1.00	1.53	
	3	66	2	97.06	0.97	2.90	

SD is blank when the number of responses scored twice is less than 10.

Table E-2. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 4

Standard	LOC	Number of Exact Matches	Number of Responses Scored Twice	Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
411	1	317	4	98.75	1.00	0.93	
	2	264	5	98.14	0.99	1.26	
	3	18	0	100.00	1.00	0.00	
413	1	262	5	98.13	0.96	3.66	
	2	237	0	100.00	1.00	0.00	
	3	78	2	97.50	0.90	5.15	
432	1	401	5	98.77	1.00	1.18	
	2	114	2	98.28	0.93	6.25	
	3	70	2	97.22	0.93	3.75	
442	1	127	3	97.69	0.99	2.10	
	2	431	5	98.85	0.99	1.82	
	3	22	0	100.00	1.00	0.00	
453	1	89	2	97.80	0.94	5.25	
	2	431	14	96.85	0.99	1.22	0.77
	3	61	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-3. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 5

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
512	1	351	4	98.87	1.00	1.80	
	2	208	4	98.11	0.97	2.95	
	3	42	2	95.45	0.92	3.75	
523	1	222	4	98.23	0.96	4.58	
	2	361	9	97.57	0.96	2.49	
	3	18	0	100.00	1.00	0.00	
533	1	138	4	97.18	0.97	4.18	
	2	429	6	98.62	0.98	2.50	
	3	25	1	96.15	0.62	4.00	
541	1	176	1	99.44	1.00	2.00	
	2	382	6	98.45	0.99	2.02	
	3	49	1	98.00	0.97	2.50	
552	1	283	2	99.30	1.00	1.75	
	2	268	5	98.17	0.98	2.32	
	3	49	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-4. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 6

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
611	1	216	3	98.63	0.98	4.03	
	2	316	3	99.06	0.99	2.37	
	3	59	2	96.72	0.80	5.00	
621	1	231	3	98.72	0.99	2.80	
	2	331	5	98.51	0.99	2.20	
	3	32	0	100.00	1.00	0.00	
631	1	113	5	95.76	0.93	3.34	
	2	462	9	98.09	0.99	1.64	
	3	12	0	100.00	1.00	0.00	
641	1	278	5	98.23	0.99	2.04	
	2	279	3	98.94	0.98	2.40	
	3	26	1	96.30	0.21	10.00	
651	1	219	2	99.10	1.00	1.30	
	2	274	6	97.86	0.99	1.32	
	3	97	1	98.98	1.00	1.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-5. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 7

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
713	1	437	5	98.87	0.99	2.72	
	2	117	2	98.32	0.84	7.50	
	3	67	1	98.53	0.97	3.30	
724	1	366	6	98.39	0.96	4.18	
	2	120	1	99.17	0.98	3.30	
	3	143	0	100.00	1.00	0.00	
732	1	390	5	98.73	0.99	2.84	
	2	187	4	97.91	0.88	6.00	
	3	38	0	100.00	1.00	0.00	
741	1	221	2	99.10	0.99	2.55	
	2	361	5	98.63	0.98	2.62	
	3	42	1	97.67	0.79	10.00	
753	1	315	4	98.75	0.97	3.60	
	2	55	0	100.00	1.00	0.00	
	3	260	6	97.74	0.97	2.05	

SD is blank when the number of responses scored twice is less than 10.

Table E-6. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts Grade 8

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
822	1	381	5	98.70	0.99	2.76	
	2	201	1	99.50	1.00	2.50	
	3	50	1	98.04	0.90	5.90	
823	1	407	5	98.79	0.97	5.10	
	2	131	2	98.50	0.97	4.15	
	3	80	2	97.56	0.96	3.50	
833	1	332	4	98.81	0.97	5.45	
	2	247	2	99.20	1.00	1.70	
	3	48	0	100.00	1.00	0.00	
842	1	288	4	98.63	0.99	2.23	
	2	260	1	99.62	1.00	2.00	
	3	63	0	100.00	1.00	0.00	
852	1	364	2	99.45	1.00	0.90	
	2	170	7	96.05	0.95	2.74	
	3	82	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-7. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — English Language Arts High School

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
911	1	127	0	100.00	1.00	0.00	
	2	331	2	99.40	1.00	0.10	
	3	98	3	97.03	0.87	6.33	
921	1	251	5	98.05	0.98	2.52	
	2	243	3	98.78	0.98	3.43	
	3	76	0	100.00	1.00	0.00	
931	1	154	3	98.09	0.95	3.90	
	2	342	5	98.56	0.99	1.80	
	3	74	0	100.00	1.00	0.00	
942	1	374	7	98.16	0.98	3.21	
	2	88	0	100.00	1.00	0.00	
	3	87	3	96.67	0.96	2.20	
951	1	204	1	99.51	1.00	1.00	
	2	241	6	97.57	0.99	1.25	
	3	118	4	96.72	0.97	1.90	

SD is blank when the number of responses scored twice is less than 10.

Table E-8. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 3

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
301	1	331	3	99.10	1.00	2.67	
	2	101	2	98.06	0.99	1.80	
	3	84	1	98.82	0.82	10.00	
302	1	162	1	99.39	1.00	0.30	
	2	239	4	98.35	0.94	4.00	
	3	111	0	100.00	1.00	0.00	
303	1	284	0	100.00	1.00	0.00	
	2	213	5	97.71	0.99	1.32	
	3	11	0	100.00	1.00	0.00	
304	1	134	1	99.26	1.00	2.00	
	2	355	4	98.89	0.99	2.38	
	3	10	0	100.00	1.00	0.00	
305	1	104	4	96.30	0.99	1.63	
	2	283	3	98.95	1.00	1.07	
	3	96	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-9. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 4

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
401	1	291	5	98.31	0.99	2.06	
	2	194	4	97.98	0.93	4.93	
	3	104	0	100.00	1.00	0.00	
402	1	323	2	99.38	1.00	1.25	
	2	58	1	98.31	1.00	1.20	
	3	209	5	97.66	0.91	3.72	
403	1	422	6	98.60	0.99	2.55	
	2	155	2	98.73	1.00	0.70	
	3	17	0	100.00	1.00	0.00	
404	1	322	2	99.38	1.00	2.90	
	2	216	5	97.74	0.99	1.60	
	3	49	2	96.08	0.93	2.90	
405	1	153	2	98.71	1.00	0.75	
	2	290	12	96.03	0.94	2.14	2.71
	3	140	1	99.29	0.95	6.70	

SD is blank when the number of responses scored twice is less than 10.

Table E-10. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 5

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
501	1	245	7	97.22	0.93	4.69	
	2	185	5	97.37	0.98	2.30	
	3	158	5	96.93	0.86	5.60	
502	1	378	7	98.18	0.98	3.00	
	2	147	10	93.63	0.93	2.58	2.34
	3	78	0	100.00	1.00	0.00	
503	1	456	8	98.28	0.96	4.43	
	2	96	1	98.97	1.00	1.50	
	3	33	0	100.00	1.00	0.00	
504	1	342	4	98.84	0.99	2.28	
	2	214	4	98.17	0.99	2.13	
	3	53	2	96.36	1.00	0.15	
505	1	128	1	99.22	1.00	0.60	
	2	424	5	98.83	1.00	0.72	
	3	54	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-11. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 6

Standard	LOC	Number of Exact Matches	Number of Responses Scored Twice	Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
605	1	304	7	97.75	0.97	3.43	
	2	222	3	98.67	0.87	7.33	
	3	54	0	100.00	1.00	0.00	
606	1	211	1	99.53	0.99	4.00	
	2	308	15	95.36	0.97	1.43	1.22
	3	66	3	95.65	0.97	2.17	
607	1	263	8	97.05	0.99	1.83	
	2	299	12	96.14	0.92	2.98	2.87
	3	20	0	100.00	1.00	0.00	
608	1	203	4	98.07	0.99	2.40	
	2	353	14	96.19	0.98	1.79	0.92
	3	28	0	100.00	1.00	0.00	
618	1	340	6	98.27	0.98	3.38	
	2	132	1	99.25	0.99	2.50	
	3	121	2	98.37	0.99	1.60	

SD is blank when the number of responses scored twice is less than 10.

Table E-12. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 7

Standard	LOC	Number of Exact Matches	Number of Responses Scored Twice	Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
705	1	87	3	96.67	0.99	3.00	
	2	443	4	99.11	1.00	1.95	
	3	97	1	98.98	1.00	1.00	
706	1	269	6	97.82	0.98	2.58	
	2	212	4	98.15	0.98	2.68	
	3	130	2	98.48	1.00	1.05	
707	1	465	12	97.48	0.97	2.79	2.63
	2	125	3	97.66	0.99	1.67	
	3	38	0	100.00	1.00	0.00	
708	1	332	6	98.22	0.99	2.33	
	2	59	1	98.33	0.99	1.60	
	3	236	11	95.55	0.98	1.30	0.79
710	1	223	8	96.54	0.97	2.41	
	2	273	8	97.15	0.98	2.04	
	3	126	2	98.44	0.98	2.10	

SD is blank when the number of responses scored twice is less than 10.

Table E-13. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics Grade 8

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
805	1	342	11	96.88	0.98	1.90	1.65
	2	228	3	98.70	0.98	2.50	
	3	56	2	96.55	0.97	1.85	
808	1	414	4	99.04	1.00	2.15	
	2	133	1	99.25	0.95	10.00	
	3	84	0	100.00	1.00	0.00	
809	1	337	7	97.97	0.98	1.93	
	2	239	3	98.76	0.97	4.60	
	3	44	2	95.65	0.62	3.85	
810	1	326	3	99.09	1.00	0.57	
	2	144	5	96.64	0.98	1.50	
	3	157	1	99.37	0.99	5.00	
818	1	527	7	98.69	1.00	1.93	
	2	71	1	98.61	0.83	10.00	
	3	27	1	96.43	0.39	10.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-14. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Mathematics High School

Standard	LOC	Number of		Percent Exact	Correlation	Mean Absolute Difference	Standard Deviation
		Exact Matches	Responses Scored Twice				
911	1	258	3	98.85	1.00	1.57	
	2	155	1	99.36	1.00	0.90	
	3	152	3	98.06	0.99	1.67	
912	1	110	5	95.65	0.97	2.52	
	2	381	8	97.94	0.99	2.44	
	3	43	2	95.56	0.30	10.00	
913	1	228	7	97.02	1.00	1.04	
	2	162	5	97.01	0.98	3.00	
	3	158	1	99.37	0.97	5.00	
914	1	160	3	98.16	0.98	3.97	
	2	252	11	95.82	0.98	2.34	1.90
	3	141	5	96.58	0.91	3.86	
915	1	277	4	98.58	0.95	6.18	
	2	76	0	100.00	1.00	0.00	
	3	193	5	97.47	0.93	4.46	

SD is blank when the number of responses scored twice is less than 10.

Table E-15. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Science Grade 4

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
411	1	283	3	98.95	1.00	2.23	
	2	244	4	98.39	0.96	4.55	
	3	58	1	98.31	0.79	6.00	
422	1	254	2	99.22	0.94	10.00	
	2	269	5	98.18	0.86	4.82	
	3	65	1	98.48	0.96	3.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-16. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Science Grade 8

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
813	1	257	2	99.23	0.99	3.75	
	2	259	4	98.48	1.00	1.08	
	3	117	0	100.00	1.00	0.00	
832	1	108	2	98.18	0.92	6.65	
	2	443	14	96.94	0.97	2.36	1.90
	3	82	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-17. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Science High School

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
921	1	193	2	98.97	1.00	2.00	
	2	153	5	96.84	0.98	2.00	
	3	219	6	97.33	0.97	2.58	
931	1	113	5	95.76	0.98	2.86	
	2	397	6	98.51	0.97	3.10	
	3	62	0	100.00	1.00	0.00	

SD is blank when the number of responses scored twice is less than 10.

Table E-18. 2014–15 NYSAA: Interrater Consistency Statistics for S/LOC Items — Social Studies High School

Standard	*LOC*	*Number of*		*Percent Exact*	*Correlation*	*Mean Absolute Difference*	*Standard Deviation*
		Exact Matches	*Responses Scored Twice*				
911	1	128	0	100.00	1.00	0.00	
	2	343	3	99.13	0.99	2.47	
	3	107	3	97.27	0.94	1.50	
921	1	90	0	100.00	1.00	0.00	
	2	387	5	98.72	0.99	2.44	
	3	85	2	97.70	0.85	3.75	

SD is blank when the number of responses scored twice is less than 10.

APPENDIX F—PERFORMANCE LEVEL DISTRIBUTIONS

Table F-1. 2014–15 NYSAA: Performance Level Distributions by Content Area and Grade

Content Area	*Grade*	*Performance Level*[1]	*Percent at Level* *2014–15*
English Language Arts	3	21	5.21
		22	14.21
		23	63.72
		24	16.87
	4	21	2.70
		22	11.66
		23	72.80
		24	12.84
	5	21	2.85
		22	9.53
		23	69.94
		24	17.68
	6	21	4.08
		22	10.80
		23	73.29
		24	11.83
	7	21	6.19
		22	11.40
		23	62.09
		24	20.32
	8	21	9.01
		22	9.89
		23	59.09
		24	22.01
	High School	21	5.86
		22	17.24
		23	60.00
		24	16.90
Mathematics	3	21	8.12
		22	13.47
		23	62.35
		24	16.06
	4	21	7.54
		22	14.52
		23	58.27
		24	19.67
	5	21	5.24
		22	10.64
		23	66.67
		24	17.45
	6	21	5.04
		22	16.46
		23	53.19
		24	25.30
	7	21	4.67
		22	15.01

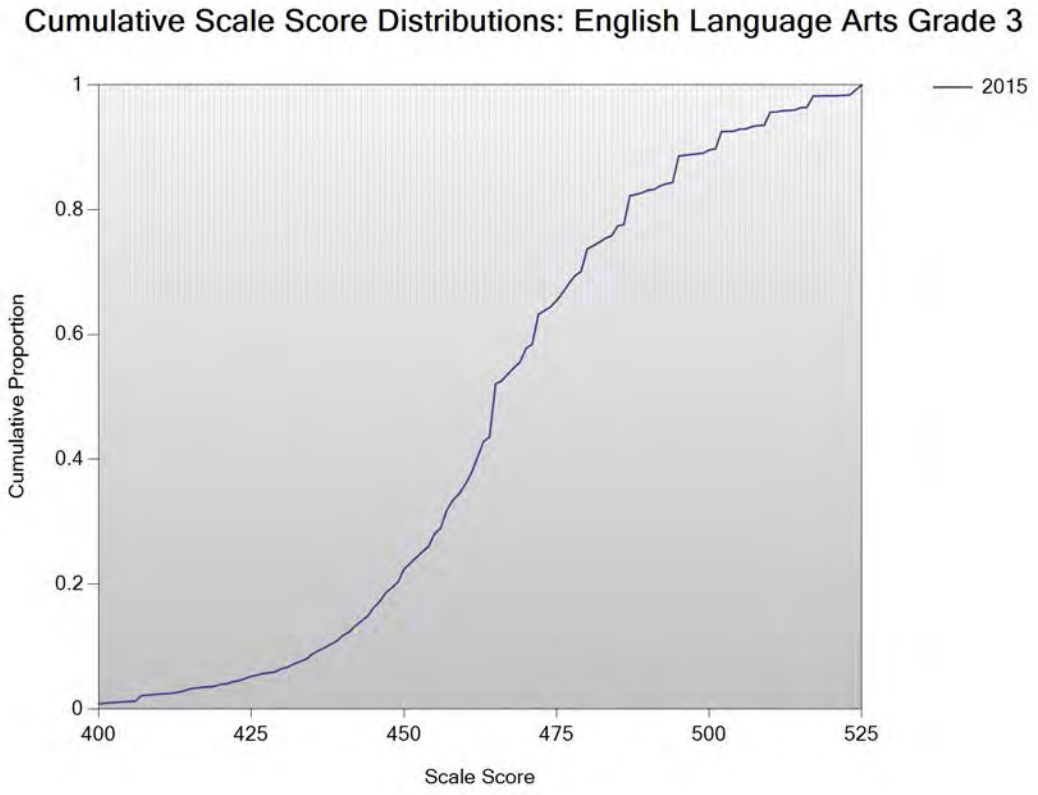
continued

Content Area	*Grade*	*Performance Level*[1]	*Percent at Level*
			2014–15
Mathematics	7	23	63.14
		24	17.18
	8	21	8.08
		22	15.21
		23	59.99
		24	16.72
	High School	21	5.51
		22	13.37
		23	60.32
		24	20.81
Science	4	21	3.89
		22	12.02
		23	70.50
		24	13.58
	8	21	3.13
		22	12.24
		23	69.57
		24	15.05
	High School	21	3.07
		22	14.02
		23	56.32
		24	26.59
Social Studies	High School	21	4.57
		22	10.60
		23	66.39
		24	18.43

[1]21 = Not Meeting Learning Standards, 22 = Partially Meeting Learning Standards, 23 = Meeting Learning Standards, 24 = Meeting Learning Standards with Distinction

APPENDIX G—CUMULATIVE DISTRIBUTION GRAPHS

Figure G-1. 2014–15 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 3 Bottom: English Language Arts Grade 4



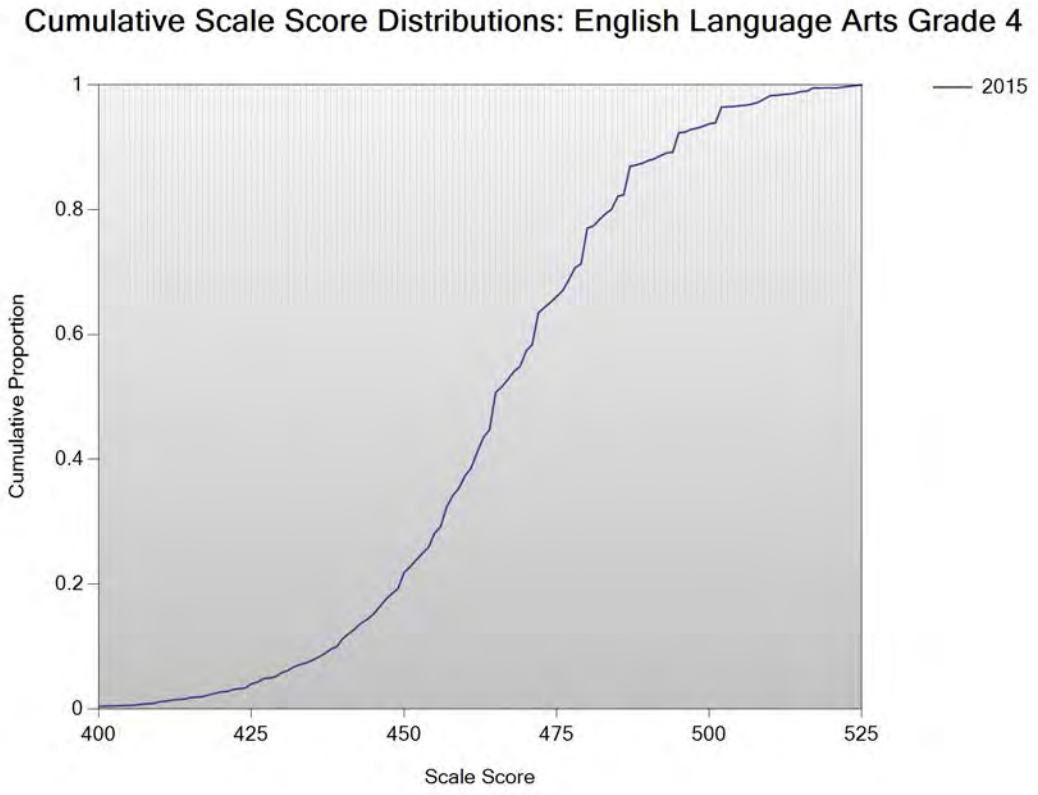
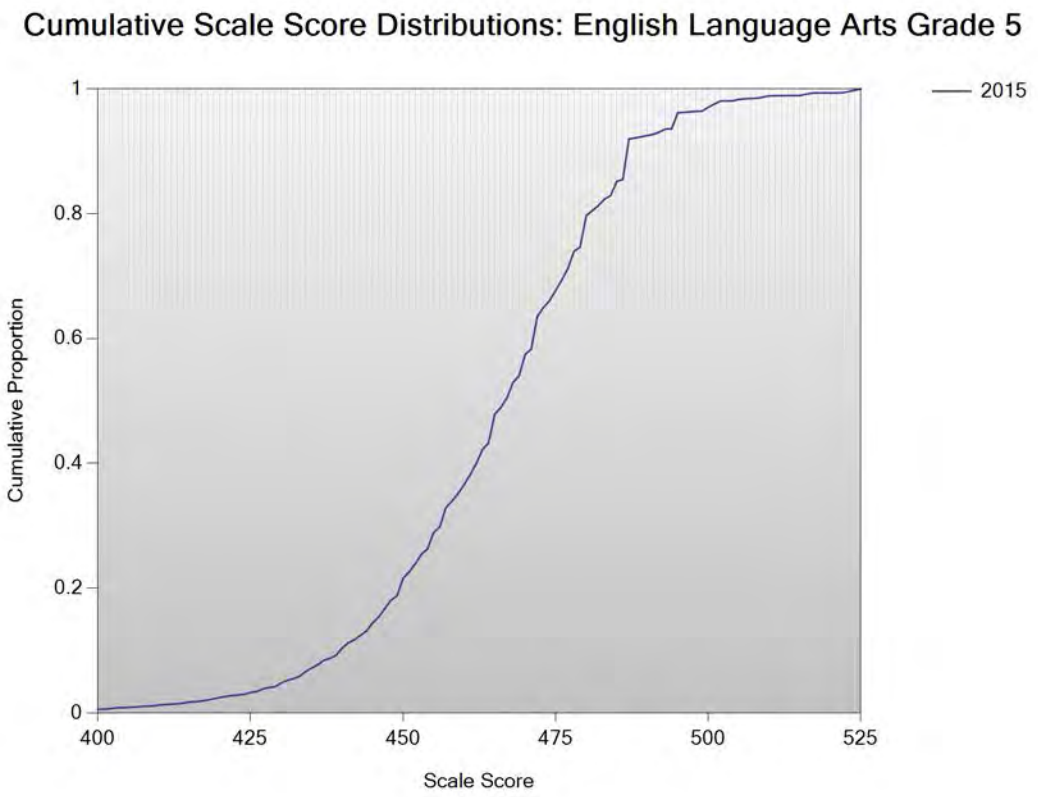


Figure G-2. 2014–15 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 5 Bottom: English Language Arts Grade 6



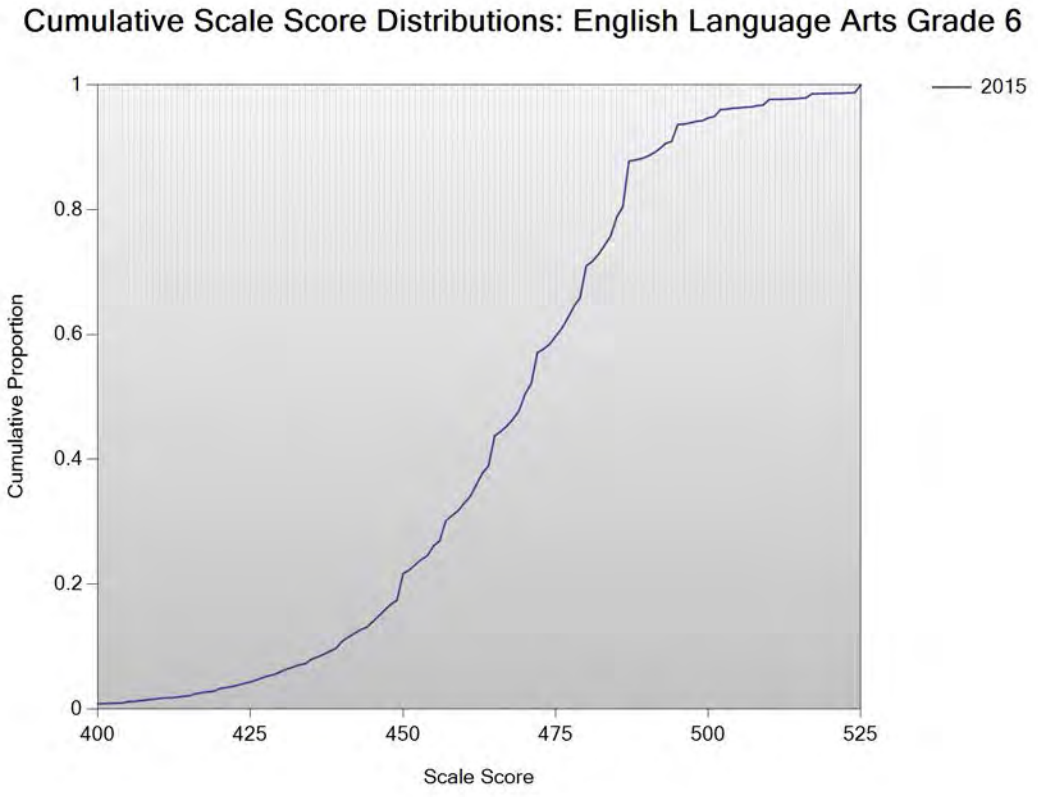
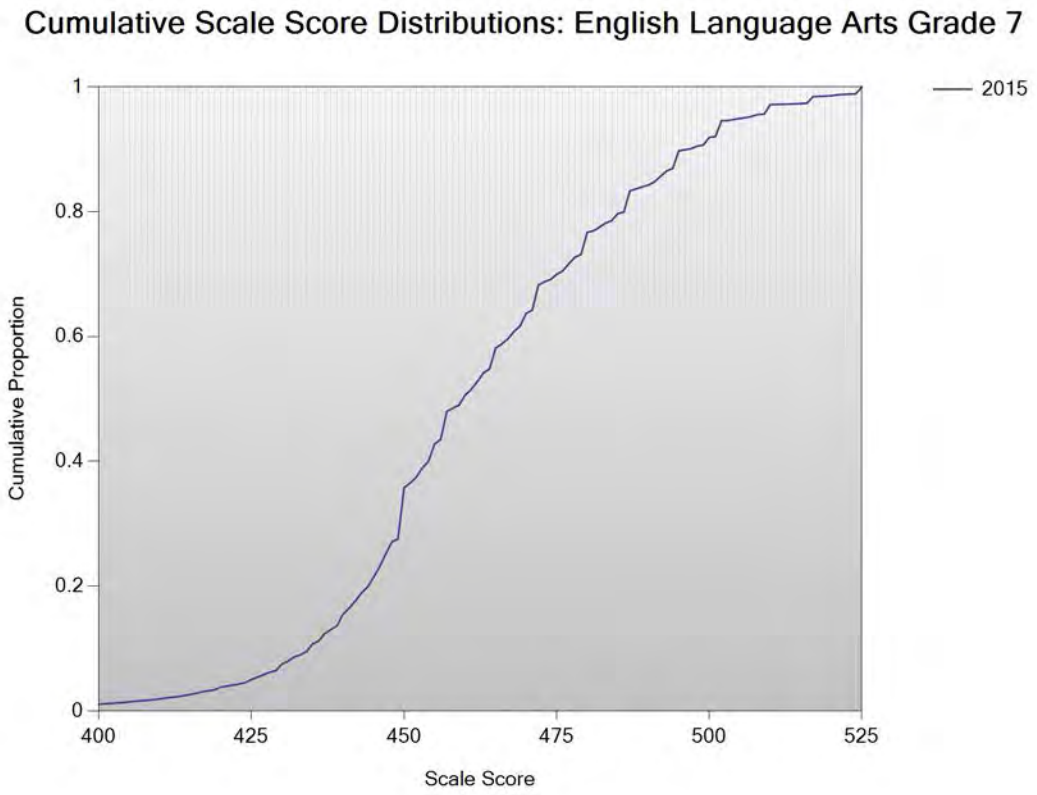


Figure G-3. 2014–15 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 7 Bottom: English Language Arts Grade 8



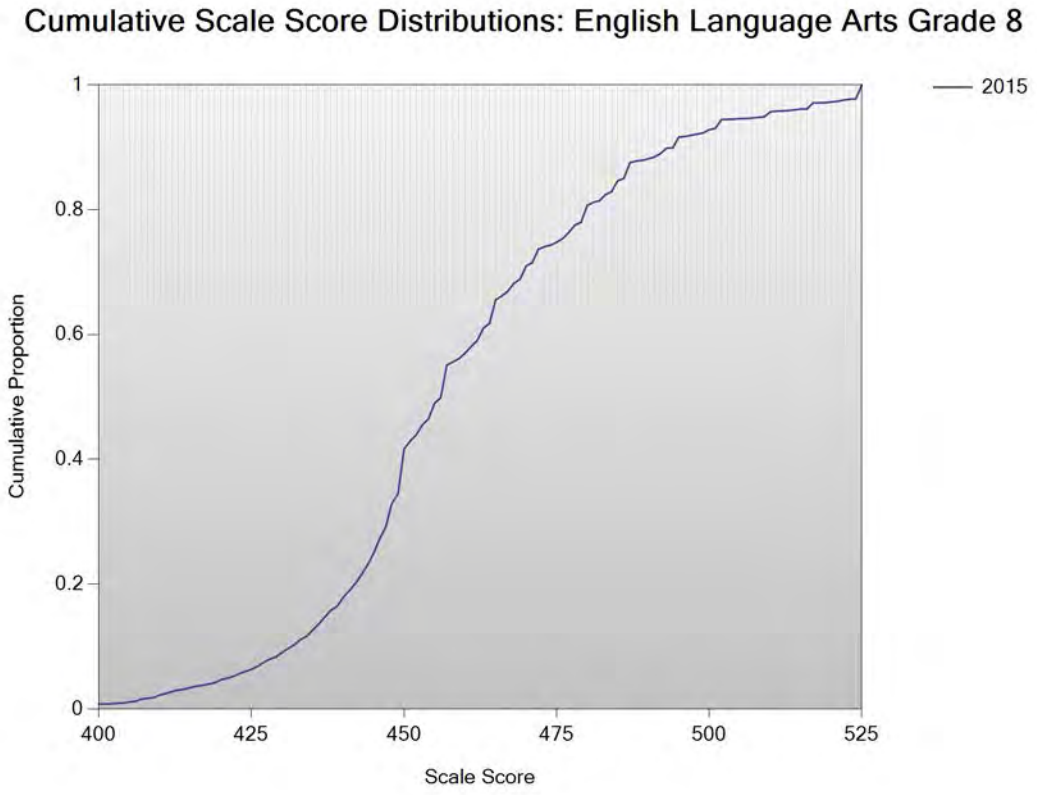
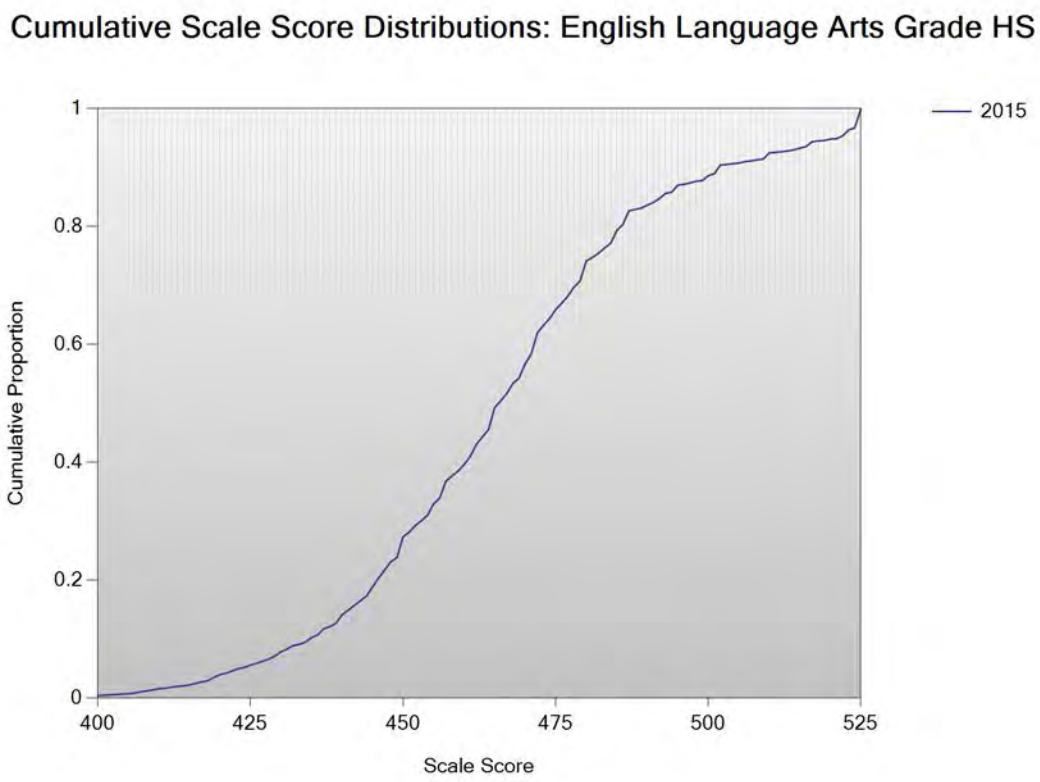


Figure G-4. 2014–15 NYSAA: Cumulative Score Distributions
Top: English Language Arts High School Bottom: Mathematics Grade 3



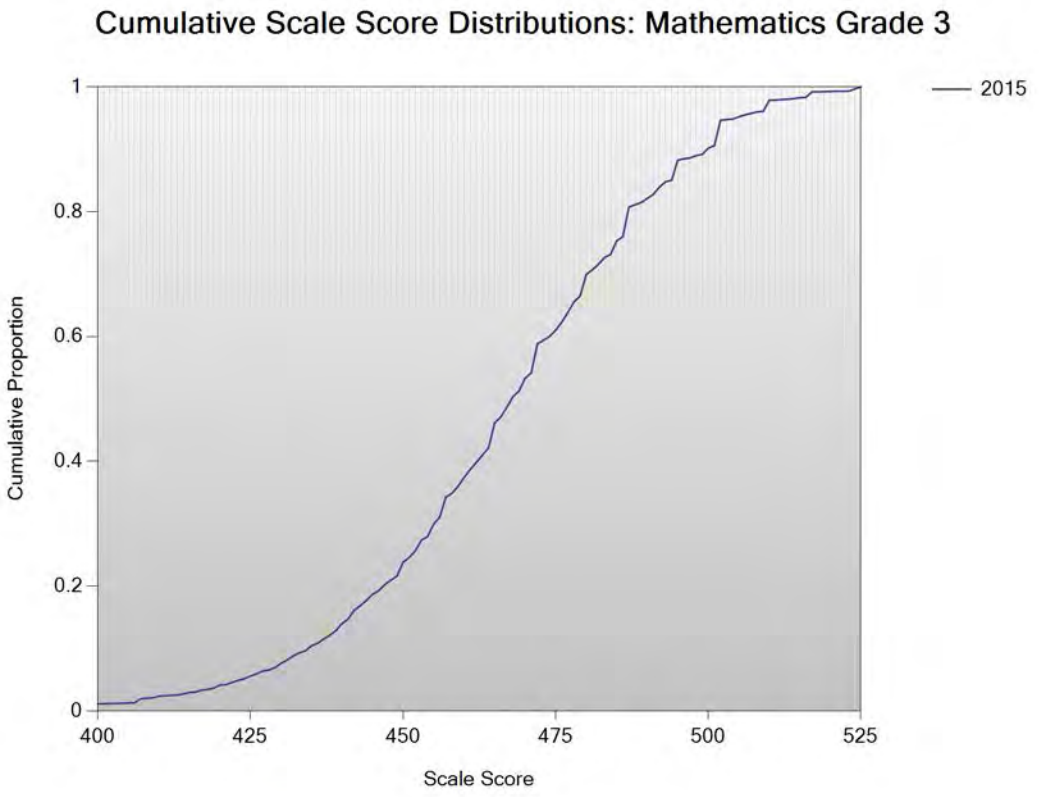
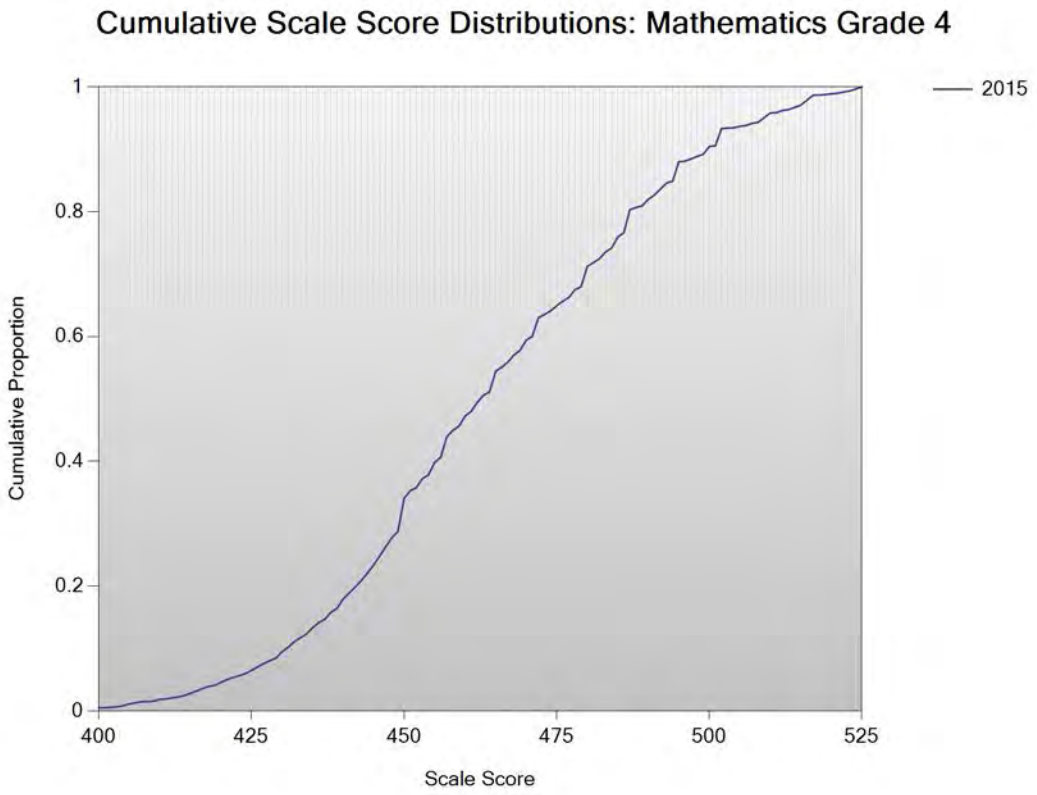


Figure G-5. 2014–15 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 4 Bottom: Mathematics Grade 5



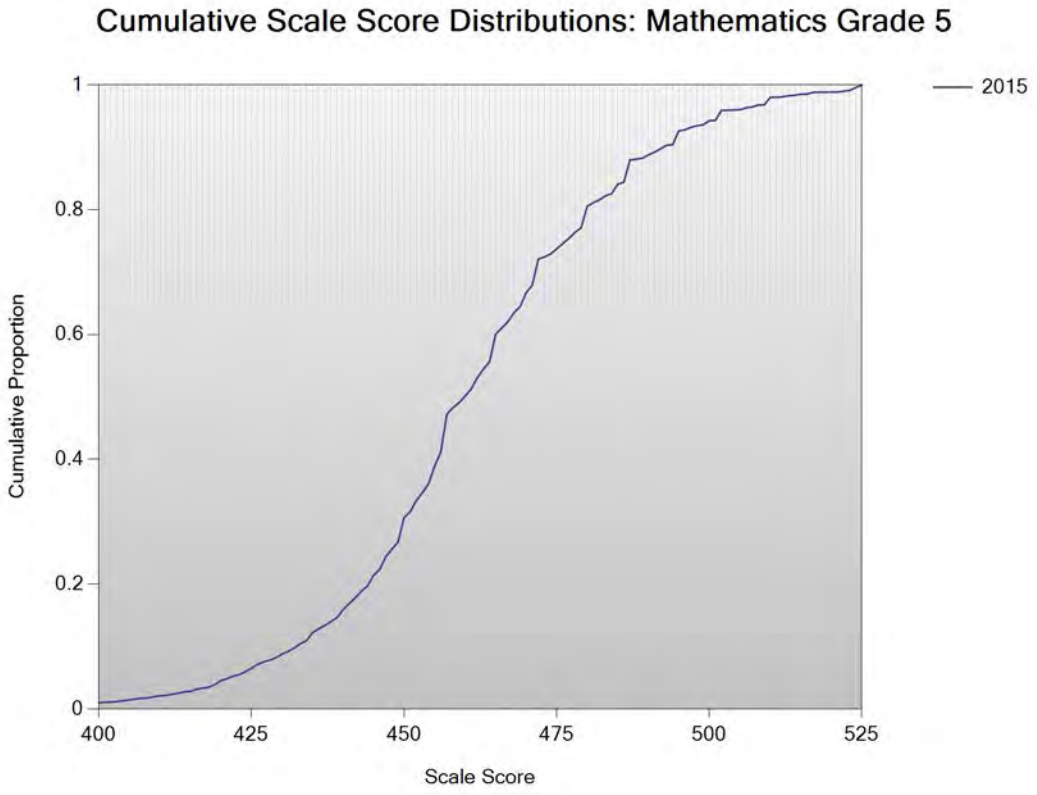
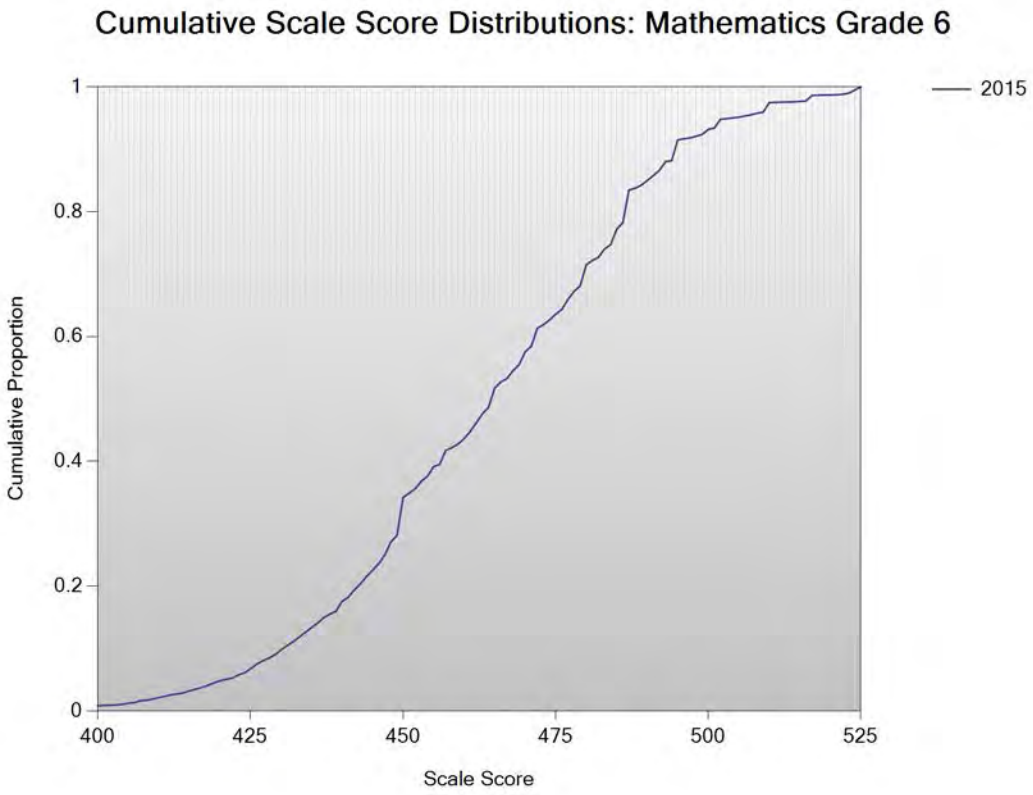


Figure G-6. 2014–15 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 6 Bottom: Mathematics Grade 7



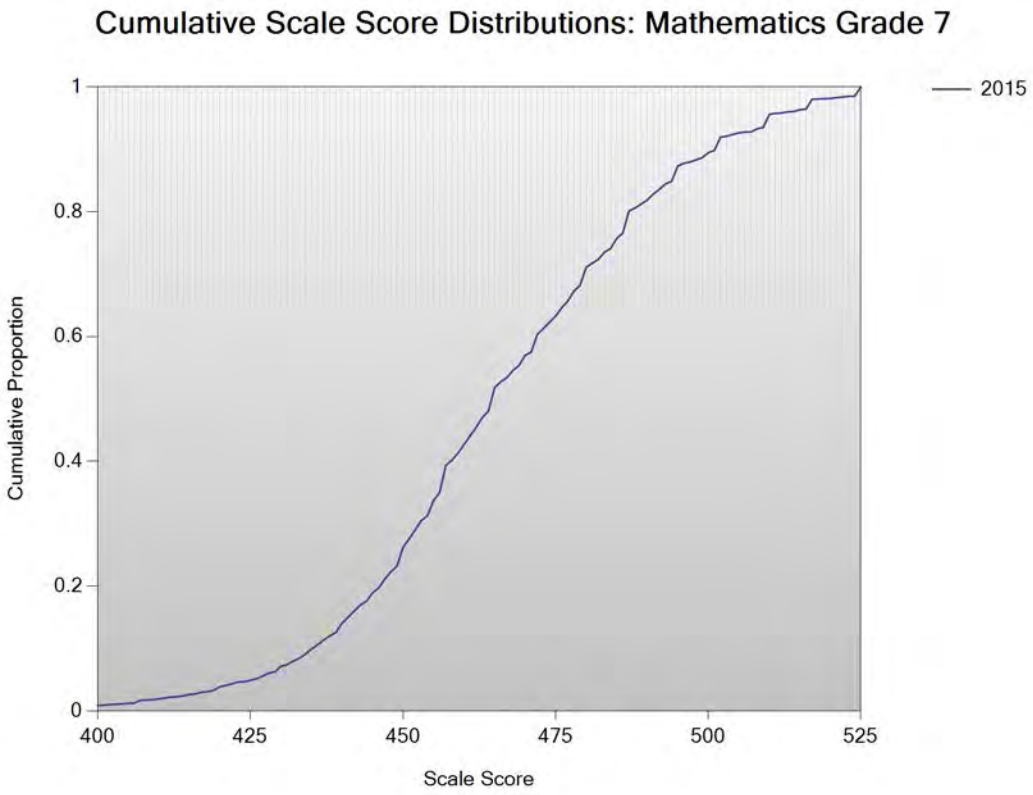
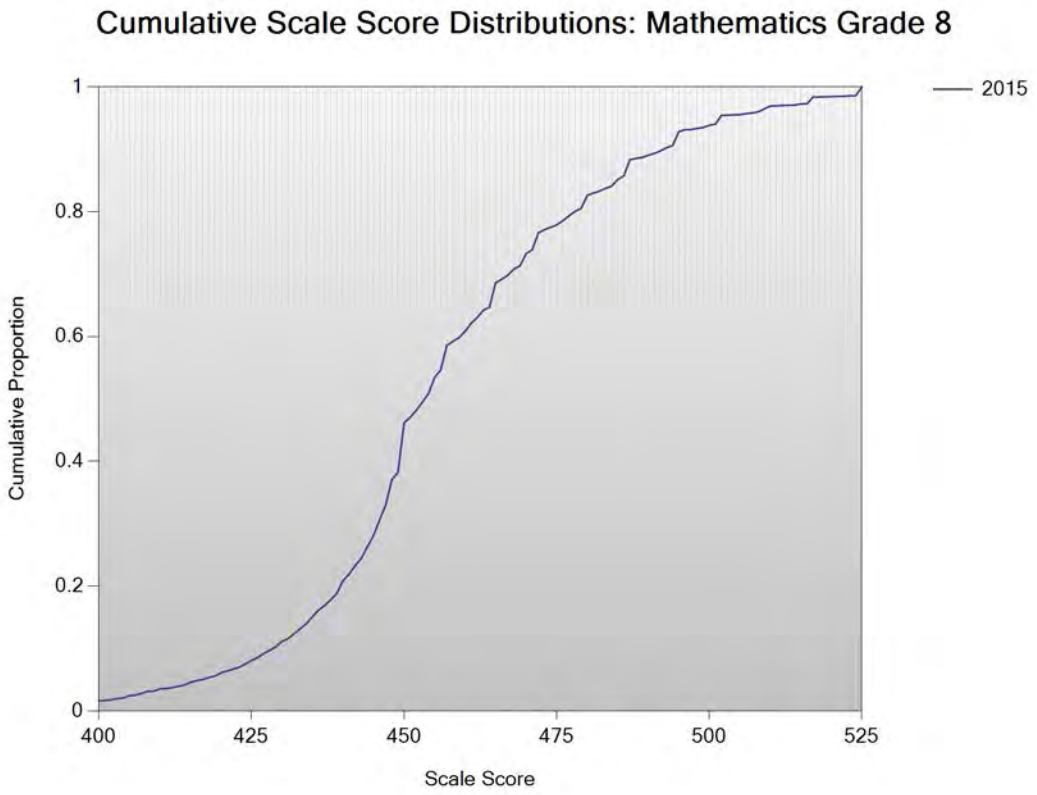


Figure G-7. 2014–15 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 8 Bottom: Mathematics High School



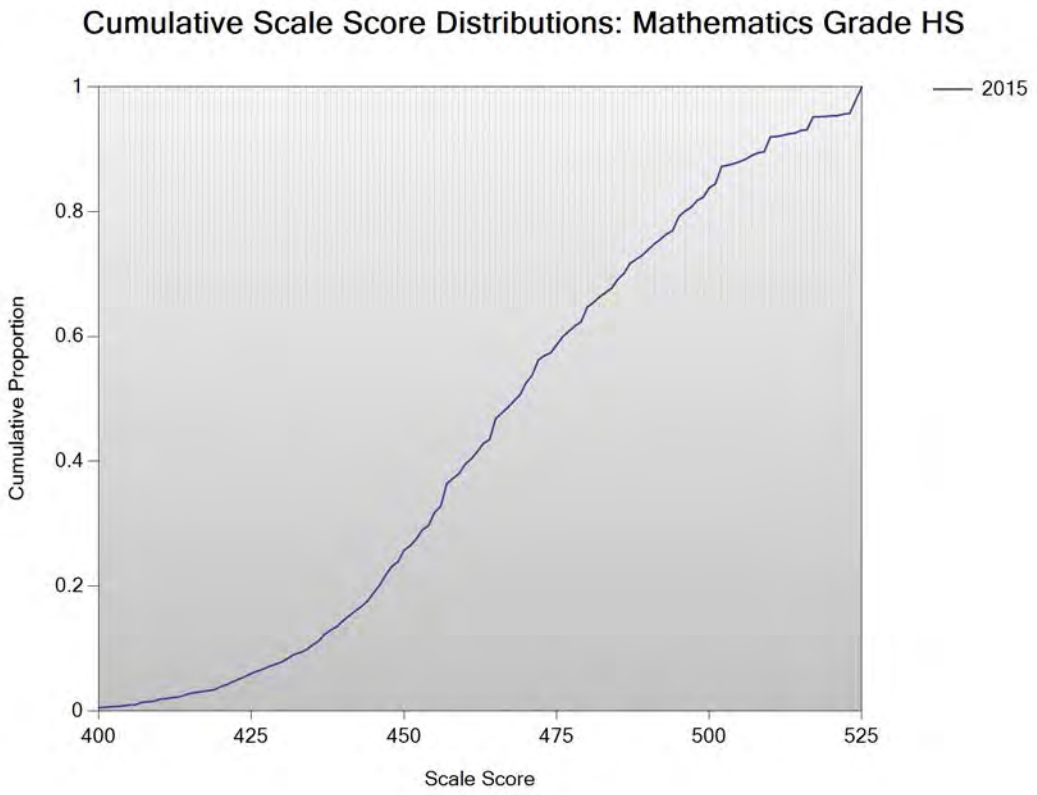
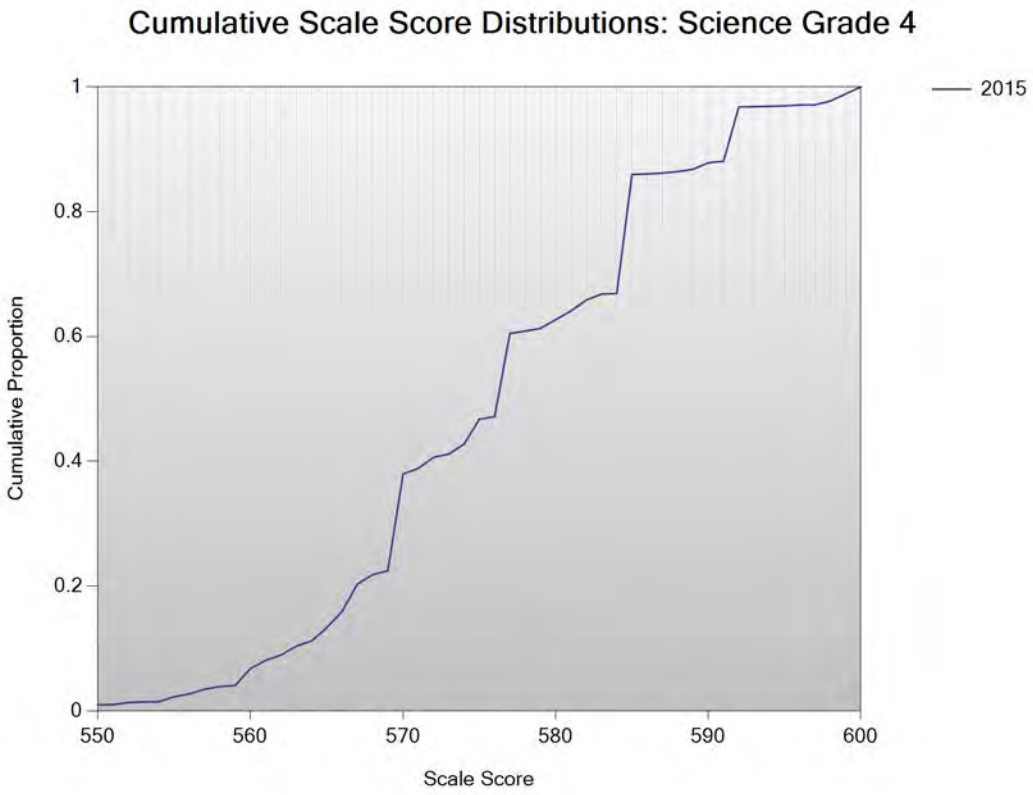


Figure G-8. 2014–15 NYSAA: Cumulative Score Distributions
Top: Science Grade 4 Bottom: Science Grade 8



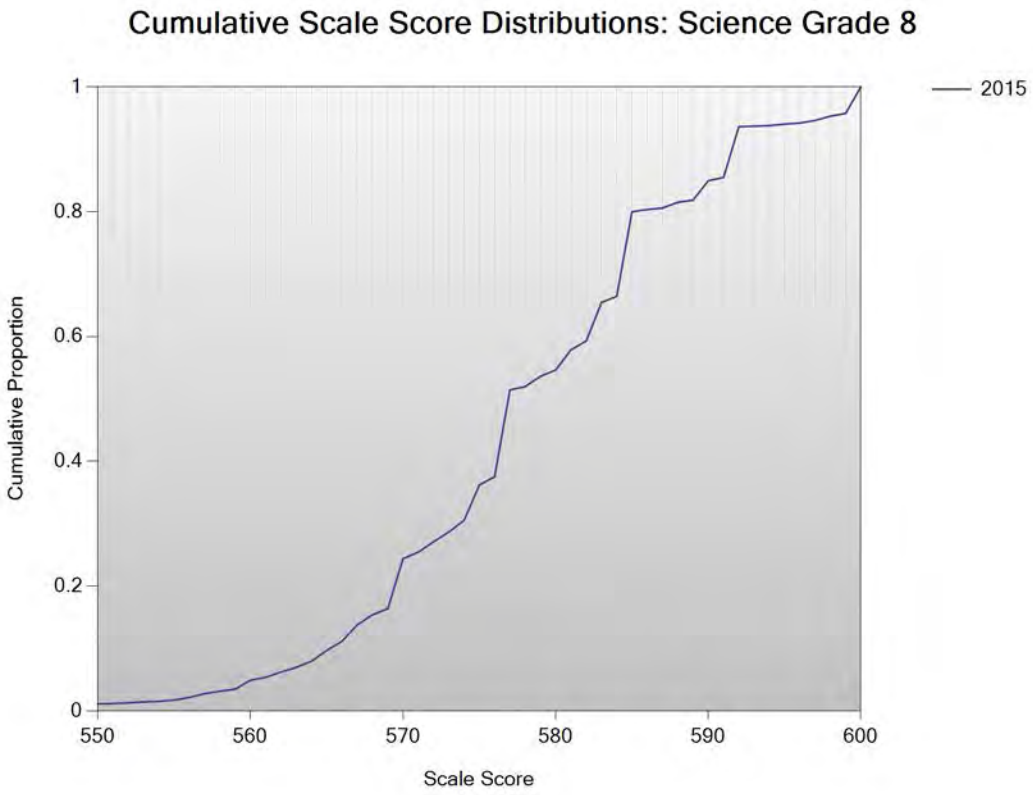
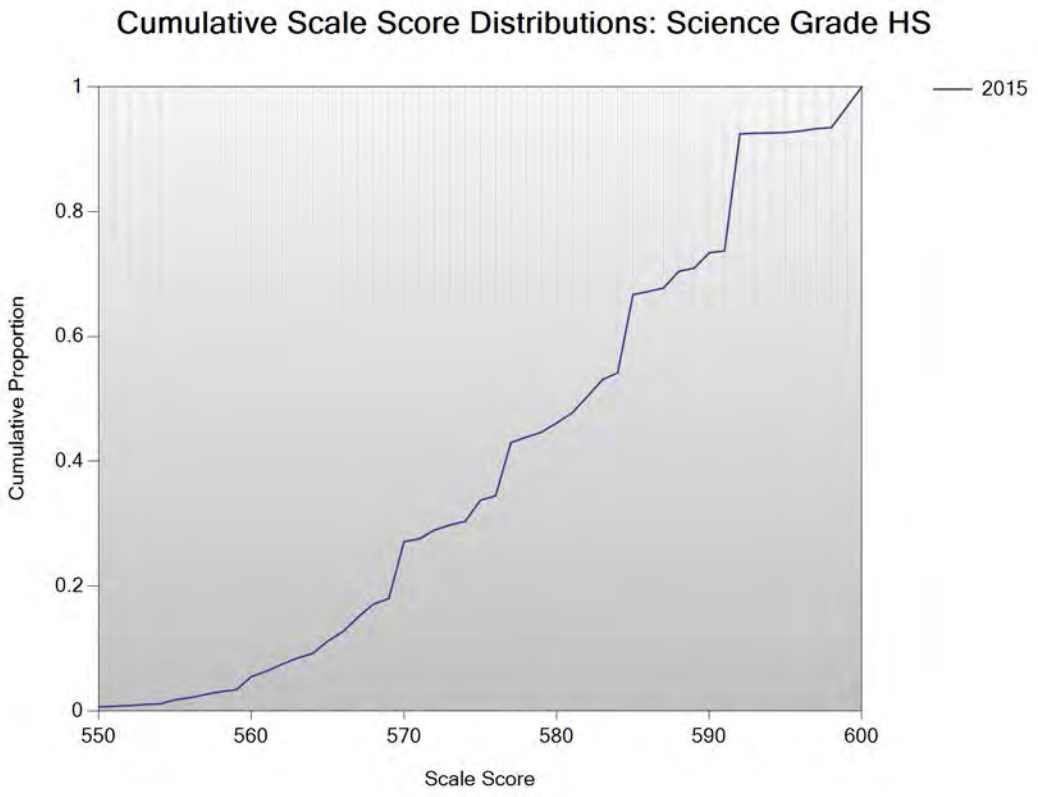
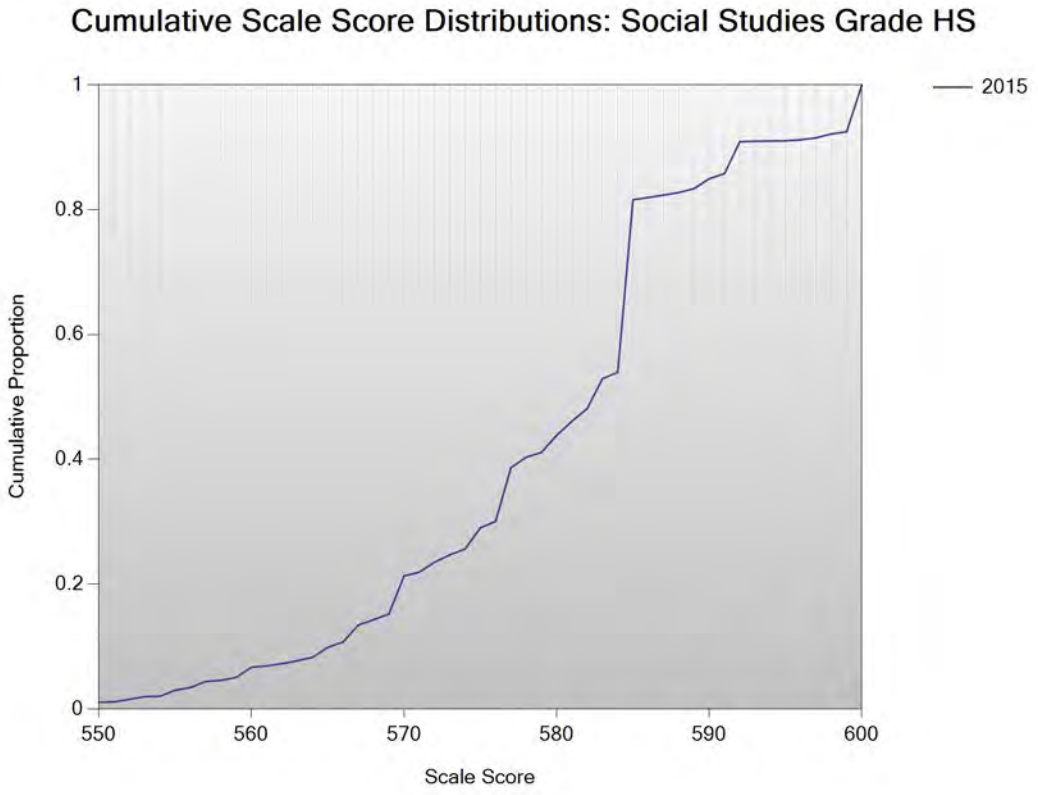


Figure G-9. 2014–15 NYSAA: Cumulative Score Distributions
Top: Science High School Bottom: Social Studies High School





APPENDIX H—CLASSICAL ITEM ANALYSIS

Table H-1. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 3

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
314	1	100	374	74.21	34.96	0.67	0.74
	2	100	2,022	87.35	23.96	0.49	0.87
	3	100	311	92.44	17.81	0.28	0.92
322	1	100	1,942	77.14	31.26	0.60	0.77
	2	100	601	86.27	22.99	0.39	0.86
	3	100	158	93.51	17.56	0.44	0.94
331	1	100	1,492	83.37	27.95	0.69	0.83
	2	100	788	91.49	16.86	0.40	0.91
	3	100	447	92.94	17.90	0.28	0.93
341	1	100	675	74.55	33.99	0.72	0.75
	2	100	1,536	87.33	26.46	0.34	0.87
	3	100	505	93.61	15.40	0.40	0.94
351	1	100	1,370	76.66	28.78	0.69	0.77
	2	100	944	87.06	24.81	0.36	0.87
	3	100	412	92.21	17.91	0.29	0.92

Table H-2. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 4

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
411	1	100	1,662	82.97	27.52	0.54	0.83
	2	100	1,172	91.74	16.42	0.36	0.92
	3	100	112	95.78	14.78	0.22	0.96
413	1	100	1,373	84.10	27.80	0.60	0.84
	2	100	1,109	91.62	18.06	0.42	0.92
	3	100	437	88.17	19.23	0.19	0.88
432	1	100	2,002	77.65	25.66	0.54	0.78
	2	100	573	86.23	23.04	0.43	0.86
	3	100	354	93.78	15.91	0.20	0.94
442	1	100	766	80.55	27.42	0.52	0.81
	2	100	2,016	88.82	20.11	0.43	0.89
	3	100	140	91.56	14.43	0.08	0.92
453	1	100	500	74.00	33.29	0.66	0.74
	2	100	2,108	85.52	20.45	0.42	0.86
	3	100	321	92.73	15.98	0.32	0.93

Table H-3. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 5

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
512	1	100	1,854	80.47	25.52	0.53	0.80
	2	100	1,014	88.04	18.41	0.40	0.88
	3	100	214	86.04	23.55	0.41	0.86
523	1	100	1,209	80.51	29.09	0.61	0.81
	2	100	1,804	88.70	17.80	0.43	0.89
	3	100	77	96.73	11.94	0.16	0.97
533	1	100	790	78.48	31.92	0.64	0.78
	2	100	2,200	88.60	20.87	0.39	0.89
	3	100	87	92.75	16.60	0.28	0.93
541	1	100	930	84.10	29.13	0.55	0.84
	2	100	1,939	87.05	18.50	0.42	0.87
	3	100	225	94.57	16.05	0.19	0.95
552	1	100	1,458	79.21	27.49	0.58	0.79
	2	100	1,357	88.08	20.26	0.37	0.88
	3	100	246	86.61	24.69	0.33	0.87

Table H-4. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 6

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
611	1	100	1,236	79.75	32.72	0.55	0.80
	2	100	1,640	91.65	19.48	0.40	0.92
	3	100	289	94.50	14.88	0.17	0.94
621	1	100	1,369	85.05	30.77	0.57	0.85
	2	100	1,626	91.23	17.30	0.32	0.91
	3	100	174	98.04	8.05	0.36	0.98
631	1	100	702	78.33	31.87	0.64	0.78
	2	100	2,381	88.97	16.90	0.42	0.89
	3	100	76	96.21	13.51	0.36	0.96
641	1	100	1,609	85.53	27.21	0.59	0.86
	2	100	1,388	90.54	16.96	0.39	0.91
	3	100	153	96.92	11.68	0.22	0.97
651	1	100	1,255	78.19	31.49	0.60	0.78
	2	100	1,369	88.38	16.59	0.37	0.88
	3	100	556	90.85	16.48	0.35	0.91

Table H-5. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 7

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
713	1	100	2,387	86.70	26.98	0.45	0.87
	2	100	566	89.94	19.21	0.31	0.90
	3	100	345	93.08	15.76	0.39	0.93
724	1	100	2,114	81.88	26.47	0.53	0.82
	2	100	564	89.48	18.60	0.32	0.89
	3	100	630	92.18	15.25	0.29	0.92
732	1	100	2,170	84.86	27.18	0.48	0.85
	2	100	966	88.30	20.91	0.32	0.88
	3	100	157	96.25	9.74	0.07	0.96
741	1	100	1,231	82.51	31.85	0.58	0.83
	2	100	1,869	88.78	21.56	0.34	0.89
	3	100	209	88.60	24.04	0.38	0.89
753	1	100	1,758	80.28	26.38	0.58	0.80
	2	100	289	87.44	19.43	0.38	0.87
	3	100	1,278	91.34	15.55	0.31	0.91

Table H-6. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts Grade 8

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
822	1	100	2,069	78.92	29.50	0.54	0.79
	2	100	1,029	91.07	21.98	0.38	0.91
	3	100	242	93.74	14.25	0.27	0.94
823	1	100	2,237	85.23	28.26	0.55	0.85
	2	100	626	87.39	23.80	0.39	0.87
	3	100	459	93.38	16.17	0.39	0.93
833	1	100	1,899	82.75	31.49	0.59	0.83
	2	100	1,165	89.68	21.61	0.41	0.90
	3	100	276	95.53	11.75	0.34	0.96
842	1	100	1,694	83.74	24.08	0.52	0.84
	2	100	1,249	87.66	19.38	0.34	0.88
	3	100	351	93.13	14.94	0.37	0.93
852	1	100	2,102	81.21	27.45	0.56	0.81
	2	100	833	86.85	17.89	0.25	0.87
	3	100	427	94.08	13.99	0.35	0.94

Table H-7. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
English Language Arts High School

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
911	1	100	749	75.51	35.71	0.57	0.76
	2	100	1,587	85.49	23.18	0.38	0.85
	3	100	504	88.37	22.74	0.35	0.88
921	1	100	1,343	81.01	27.49	0.61	0.81
	2	100	1,148	84.69	24.51	0.44	0.85
	3	100	356	92.83	17.90	0.56	0.93
931	1	100	969	80.85	27.56	0.60	0.81
	2	100	1,557	91.37	17.55	0.27	0.91
	3	100	324	92.65	15.64	0.56	0.93
942	1	100	2,013	83.09	26.75	0.53	0.83
	2	100	428	87.65	18.33	0.38	0.88
	3	100	401	89.74	17.36	0.44	0.90
951	1	100	1,127	75.87	28.67	0.61	0.76
	2	100	1,211	80.89	22.39	0.25	0.81
	3	100	519	91.30	16.86	0.32	0.91

Table H-8. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 3

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
301	1	100	1,756	78.23	29.55	0.67	0.78
	2	100	542	83.28	21.86	0.55	0.83
	3	100	425	89.30	19.55	0.42	0.89
302	1	100	893	73.05	33.06	0.66	0.73
	2	100	1,243	85.16	22.66	0.44	0.85
	3	100	587	91.63	15.38	0.38	0.92
303	1	100	1,527	82.58	26.86	0.63	0.83
	2	100	1,080	87.85	19.88	0.51	0.88
	3	100	112	91.16	17.75	0.44	0.91
304	1	100	809	73.04	34.38	0.68	0.73
	2	100	1,821	86.33	22.26	0.48	0.86
	3	100	92	92.14	17.89	0.17	0.92
305	1	100	590	73.37	34.21	0.75	0.73
	2	100	1,600	85.22	23.35	0.57	0.85
	3	100	516	92.96	17.10	0.43	0.93

Table H-9. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 4

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
401	1	100	1,490	77.14	28.78	0.54	0.77
	2	100	952	85.02	24.85	0.34	0.85
	3	100	498	91.20	18.33	0.22	0.91
402	1	100	1,585	72.22	31.84	0.63	0.72
	2	100	316	80.79	24.24	0.44	0.81
	3	100	1,031	87.87	18.91	0.43	0.88
403	1	100	2,055	81.68	27.04	0.55	0.82
	2	100	804	87.60	21.79	0.34	0.88
	3	100	84	94.64	15.02	0.35	0.95
404	1	100	1,615	80.14	27.86	0.58	0.80
	2	100	1,023	88.05	18.56	0.41	0.88
	3	100	277	91.57	14.29	0.27	0.92
405	1	100	895	81.05	27.97	0.60	0.81
	2	100	1,383	85.44	22.43	0.49	0.85
	3	100	642	90.24	20.44	0.34	0.90

Table H-10. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 5

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
501	1	100	1,387	79.16	29.63	0.70	0.79
	2	100	897	84.64	24.71	0.30	0.85
	3	100	777	88.25	21.63	0.28	0.88
502	1	100	2,036	79.21	28.28	0.63	0.79
	2	100	743	85.34	23.39	0.26	0.85
	3	100	324	91.00	19.22	0.30	0.91
503	1	100	2,449	81.28	28.73	0.60	0.81
	2	100	440	86.18	24.96	0.46	0.86
	3	100	155	87.50	22.17	0.53	0.88
504	1	100	1,817	83.28	25.83	0.65	0.83
	2	100	1,043	86.64	20.76	0.31	0.87
	3	100	242	89.44	21.91	0.40	0.89
505	1	100	748	76.61	30.74	0.71	0.77
	2	100	2,088	91.09	16.72	0.37	0.91
	3	100	245	90.60	21.57	0.38	0.91

Table H-11. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 6

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
605	1	100	1,712	83.53	27.22	0.59	0.84
	2	100	1,183	89.60	21.64	0.33	0.90
	3	100	267	93.78	16.39	0.40	0.94
606	1	100	1,227	79.11	27.27	0.63	0.79
	2	100	1,531	91.08	18.47	0.33	0.91
	3	100	403	90.87	18.71	0.36	0.91
607	1	100	1,507	77.09	31.61	0.64	0.77
	2	100	1,524	89.61	21.65	0.36	0.90
	3	100	108	93.75	15.93	0.56	0.94
608	1	100	1,235	78.06	28.38	0.67	0.78
	2	100	1,782	87.09	19.41	0.39	0.87
	3	100	155	94.92	14.29	0.35	0.95
618	1	100	1,845	76.29	30.19	0.61	0.76
	2	100	646	89.27	20.63	0.50	0.89
	3	100	680	89.89	18.42	0.42	0.90

Table H-12. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 7

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
705	1	100	537	69.65	38.34	0.52	0.70
	2	100	2,353	84.90	22.68	0.49	0.85
	3	100	431	92.88	14.58	0.36	0.93
706	1	100	1,522	75.19	30.76	0.56	0.75
	2	100	1,144	85.07	24.04	0.48	0.85
	3	100	628	91.11	18.42	0.33	0.91
707	1	100	2,557	78.96	31.00	0.54	0.79
	2	100	570	83.85	24.76	0.52	0.84
	3	100	181	92.86	14.90	0.32	0.93
708	1	100	1,836	79.71	28.48	0.60	0.80
	2	100	317	87.26	20.63	0.38	0.87
	3	100	1,165	88.66	17.33	0.42	0.89
710	1	100	1,292	78.53	29.29	0.65	0.79
	2	100	1,381	83.84	22.27	0.39	0.84
	3	100	644	95.61	14.48	0.45	0.96

Table H-13. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics Grade 8

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
805	1	100	1,964	83.04	27.14	0.62	0.83
	2	100	1,102	87.67	20.45	0.35	0.88
	3	100	279	93.46	17.97	0.36	0.93
808	1	100	2,258	84.30	29.12	0.59	0.84
	2	100	713	84.16	31.31	0.47	0.84
	3	100	379	92.15	18.29	0.51	0.92
809	1	100	1,932	76.46	30.92	0.63	0.76
	2	100	1,173	88.40	21.77	0.36	0.88
	3	100	213	94.53	16.11	0.44	0.95
810	1	100	1,866	81.42	31.43	0.64	0.81
	2	100	686	90.50	19.86	0.25	0.91
	3	100	763	86.46	29.71	0.41	0.86
818	1	100	2,821	78.36	29.91	0.55	0.78
	2	100	376	92.50	18.05	0.31	0.93
	3	100	163	93.09	15.15	0.33	0.93

Table H-14. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Mathematics High School

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
911	1	100	1,357	80.16	27.21	0.58	0.80
	2	100	747	88.84	17.73	0.38	0.89
	3	100	743	89.02	21.07	0.39	0.89
912	1	100	643	68.04	36.22	0.60	0.68
	2	100	1,889	81.72	25.83	0.44	0.82
	3	100	276	95.39	13.58	0.42	0.95
913	1	100	1,260	76.44	32.22	0.63	0.76
	2	100	775	79.98	30.95	0.51	0.80
	3	100	788	90.73	20.74	0.45	0.91
914	1	100	958	76.27	32.00	0.54	0.76
	2	100	1,159	77.63	28.07	0.32	0.78
	3	100	721	91.30	21.12	0.33	0.91
915	1	100	1,494	78.94	29.42	0.54	0.79
	2	100	478	84.92	27.41	0.35	0.85
	3	100	860	88.05	21.79	0.47	0.88

Table H-15. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Science Grade 4

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
411	1	100	1,405	81.00	27.21	0.46	0.81
	2	100	1,205	89.45	21.82	0.27	0.89
	3	100	292	95.26	11.62	0.08	0.95
422	1	100	1,289	83.76	26.33	0.42	0.84
	2	100	1,309	92.97	16.12	0.24	0.93
	3	100	324	91.79	19.53	0.30	0.92

Table H-16. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Science Grade 8

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
813	1	100	1,415	82.16	28.51	0.50	0.82
	2	100	1,320	87.27	20.38	0.25	0.87
	3	100	592	93.51	15.08	0.22	0.94
832	1	100	743	79.34	29.28	0.46	0.79
	2	100	2,190	86.04	21.29	0.36	0.86
	3	100	404	93.35	14.14	0.29	0.93

Table H-17. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Science High School

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
921	1	100	1,058	81.27	26.33	0.48	0.81
	2	100	777	85.86	22.05	0.35	0.86
	3	100	1,020	91.05	19.95	0.28	0.91
931	1	100	654	78.13	29.92	0.47	0.78
	2	100	1,909	87.44	23.86	0.37	0.87
	3	100	294	94.81	14.95	0.25	0.95

Table H-18. 2014–15 NYSAA: Classical Test Theory Statistics for S/LOC Items
Social Studies High School

Standard	*LOC*	*Maximum Score*	*N*	*Mean*	*SD*	*Discrimination*	*Difficulty*
911	1	100	737	75.25	31.44	0.46	0.75
	2	100	1,662	87.54	24.34	0.34	0.88
	3	100	457	93.91	12.90	0.31	0.94
921	1	100	528	67.32	41.30	0.54	0.67
	2	100	1,924	85.64	24.90	0.38	0.86
	3	100	388	94.94	14.40	0.30	0.95

Table H-19. 2014–15 NYSAA: Classical Test Theory Statistics for Standards-Based Items

Subject	*Grade*	*Item*	*Maximum Score*	*N*	*Mean*	*SD*	*Difficulty*	*Discrimination*
English Language Arts	3	314	25	2,707	15.94	4.94	0.64	0.63
		322	25	2,701	10.56	5.68	0.42	0.71
		331	25	2,727	13.35	6.49	0.53	0.73
		341	25	2,716	15.56	6.19	0.62	0.67
		351	25	2,726	13.13	6.61	0.53	0.73
	4	411	25	2,946	12.25	5.27	0.49	0.68
		413	25	2,919	13.85	6.10	0.55	0.60
		432	25	2,929	11.41	6.29	0.46	0.61
		442	25	2,922	14.57	4.76	0.58	0.60
		453	25	2,929	15.48	5.00	0.62	0.65
	5	512	25	3,082	11.84	5.50	0.47	0.60
		523	25	3,090	13.32	4.97	0.53	0.63
		533	25	3,077	14.40	4.73	0.58	0.63
		541	25	3,094	14.46	5.01	0.58	0.61
		552	25	3,061	12.90	5.67	0.52	0.57
	6	611	25	3,165	13.98	5.85	0.56	0.66
		621	25	3,169	13.57	5.33	0.54	0.71
		631	25	3,159	14.69	4.41	0.59	0.66
		641	25	3,150	12.86	5.24	0.51	0.69
		651	25	3,180	14.33	6.38	0.57	0.67
	7	713	25	3,298	11.65	5.78	0.47	0.66
		724	25	3,308	12.65	6.80	0.51	0.64
		732	25	3,293	11.56	5.22	0.46	0.68
		741	25	3,309	13.83	5.33	0.55	0.62
		753	25	3,325	14.93	7.93	0.60	0.65
	8	822	25	3,340	11.77	5.92	0.47	0.71
		823	25	3,322	12.16	6.27	0.49	0.71
		833	25	3,340	12.48	5.92	0.50	0.76
		842	25	3,294	13.06	5.77	0.52	0.67
		852	25	3,362	12.19	6.25	0.49	0.71
	High School	911	25	2,840	15.19	6.03	0.61	0.71
		921	25	2,847	13.30	6.09	0.53	0.79
		931	25	2,850	14.60	5.65	0.58	0.76
		942	25	2,842	11.72	6.21	0.47	0.73
		951	25	2,857	13.98	6.47	0.56	0.77
Mathematics	3	301	25	2,723	11.93	6.60	0.48	0.70
		302	25	2,723	14.92	6.67	0.60	0.74
		303	25	2,719	12.10	5.19	0.48	0.76

continued

Subject	*Grade*	*Item*	*Maximum Score*	*N*	*Mean*	*SD*	*Difficulty*	*Discrimination*
Mathematics	3	304	25	2,722	13.78	5.18	0.55	0.71
		305	25	2,706	15.71	5.98	0.63	0.69
	4	401	25	2,940	13.18	6.69	0.53	0.74
		402	25	2,932	13.95	8.13	0.56	0.67
		403	25	2,943	10.84	4.98	0.43	0.70
		404	25	2,915	12.46	5.90	0.50	0.74
		405	25	2,920	15.37	6.21	0.61	0.65
	5	501	25	3,061	14.31	7.02	0.57	0.67
		502	25	3,103	11.55	6.07	0.46	0.71
		503	25	3,044	10.08	5.06	0.40	0.70
		504	25	3,102	12.18	5.53	0.49	0.71
		505	25	3,081	15.03	5.09	0.60	0.60
	6	605	25	3,162	12.74	5.75	0.51	0.74
		606	25	3,161	14.19	5.97	0.57	0.77
		607	25	3,139	12.53	5.54	0.50	0.78
		608	25	3,172	13.34	5.29	0.53	0.80
		618	25	3,171	12.93	7.21	0.52	0.70
	7	705	25	3,321	15.61	5.32	0.62	0.61
		706	25	3,294	13.63	6.87	0.55	0.66
		707	25	3,308	10.17	5.40	0.41	0.67
		708	25	3,318	14.34	7.79	0.57	0.67
		710	25	3,317	14.44	6.59	0.58	0.62
	8	805	25	3,345	12.27	5.71	0.49	0.74
		808	25	3,350	11.81	6.07	0.47	0.72
		809	25	3,318	11.80	5.92	0.47	0.76
		810	25	3,315	13.45	7.08	0.54	0.68
		818	25	3,360	9.63	5.18	0.39	0.70
	High School	911	25	2,847	14.36	7.04	0.57	0.70
		912	25	2,808	14.51	5.66	0.58	0.63
		913	25	2,823	14.39	7.44	0.58	0.65
		914	25	2,838	14.94	6.88	0.60	0.70
		915	25	2,832	14.09	7.54	0.56	0.71
Science	4	411	25	2,902	13.22	5.96	0.53	0.56
		422	25	2,922	13.90	5.80	0.56	0.54
	8	813	25	3,327	14.27	6.38	0.57	0.51
		832	25	3,337	15.28	5.25	0.61	0.53
	High School	921	25	2,855	16.00	7.19	0.64	0.62
		931	25	2,857	15.16	5.31	0.61	0.62
Social Studies	High School	911	25	2,856	15.30	5.97	0.61	0.63
		921	25	2,840	15.48	5.76	0.62	0.66

APPENDIX I—CORRELATIONS BETWEEN STANDARDS-BASED ITEMS

Table I-1. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 3

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	314	322	2,669	0.52
	314	331	2,690	0.53
	314	341	2,681	0.57
	314	351	2,688	0.52
	322	331	2,685	0.62
	322	341	2,672	0.55
	322	351	2,682	0.66
	331	341	2,699	0.58
	331	351	2,708	0.68
	341	351	2,699	0.56
Mathematics	301	302	2,702	0.61
	301	303	2,698	0.68
	301	304	2,701	0.56
	301	305	2,686	0.55
	302	303	2,696	0.66
	302	304	2,701	0.64
	302	305	2,687	0.61
	303	304	2,696	0.63
	303	305	2,680	0.60
	304	305	2,684	0.60

Table I-2. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 4

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	411	413	2,902	0.55
	411	432	2,911	0.59
	411	442	2,906	0.50
	411	453	2,911	0.53
	413	432	2,884	0.44
	413	442	2,881	0.47
	413	453	2,883	0.50
	432	442	2,887	0.44
	432	453	2,895	0.50
	442	453	2,887	0.56
Mathematics	401	402	2,906	0.60
	401	403	2,915	0.65
	401	404	2,889	0.67
	401	405	2,895	0.57
	402	403	2,910	0.55
	402	404	2,882	0.60
	402	405	2,888	0.51
	403	404	2,892	0.64
	403	405	2,897	0.54
	404	405	2,874	0.59
Science	411	422	2,871	0.59

Table I-3. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 5

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	512	523	3,045	0.51
	512	533	3,033	0.47
	512	541	3,051	0.49
	512	552	3,020	0.47
	523	533	3,041	0.58
	523	541	3,060	0.49
	523	552	3,027	0.44
	533	541	3,048	0.50
	533	552	3,013	0.44
	541	552	3,031	0.49
Mathematics	501	502	3,037	0.58
	501	503	2,982	0.55
	501	504	3,036	0.58
	501	505	3,017	0.53
	502	503	3,023	0.67
	502	504	3,080	0.63
	502	505	3,057	0.50
	503	504	3,023	0.61
	503	505	3,003	0.48
	504	505	3,060	0.53

Table I-4. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 6

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	611	621	3,127	0.57
	611	631	3,117	0.57
	611	641	3,107	0.56
	611	651	3,136	0.55
	621	631	3,121	0.55
	621	641	3,115	0.61
	621	651	3,138	0.60
	631	641	3,106	0.55
	631	651	3,130	0.56
	641	651	3,124	0.58
Mathematics	605	606	3,115	0.66
	605	607	3,092	0.69
	605	608	3,126	0.67
	605	618	3,126	0.63
	606	607	3,088	0.70
	606	608	3,123	0.73
	606	618	3,124	0.62
	607	608	3,108	0.72
	607	618	3,099	0.64
	608	618	3,136	0.64

Table I-5. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 7

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	713	724	3,249	0.57
	713	732	3,238	0.58
	713	741	3,255	0.50
	713	753	3,267	0.51
	724	732	3,246	0.55
	724	741	3,261	0.46
	724	753	3,275	0.53
	732	741	3,249	0.55
	732	753	3,259	0.54
	741	753	3,279	0.53
Mathematics	705	706	3,262	0.48
	705	707	3,269	0.53
	705	708	3,278	0.49
	705	710	3,281	0.50
	706	707	3,241	0.58
	706	708	3,254	0.57
	706	710	3,258	0.49
	707	708	3,268	0.56
	707	710	3,263	0.51
	708	710	3,275	0.54

Table I-6. 2014–15 NYSAA: Standards-Based Item Correlations for Grade 8

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	822	823	3,272	0.61
	822	833	3,288	0.65
	822	842	3,247	0.56
	822	852	3,311	0.61
	823	833	3,273	0.66
	823	842	3,229	0.57
	823	852	3,292	0.62
	833	842	3,243	0.60
	833	852	3,306	0.65
	842	852	3,263	0.58
Mathematics	805	808	3,307	0.64
	805	809	3,279	0.65
	805	810	3,274	0.61
	805	818	3,315	0.62
	808	809	3,283	0.64
	808	810	3,281	0.57
	808	818	3,324	0.61
	809	810	3,252	0.63
	809	818	3,290	0.65
	810	818	3,288	0.55
Science	813	832	3,282	0.55

Table I-7. 2014–15 NYSAA: Standards-Based Item Correlations for High School

Subject	*Pair of Standards*		*N*	*Correlation*
English Language Arts	911	921	2,795	0.63
	911	931	2,796	0.65
	911	942	2,790	0.59
	911	951	2,802	0.63
	921	931	2,801	0.68
	921	942	2,797	0.71
	921	951	2,807	0.72
	931	942	2,797	0.63
	931	951	2,815	0.69
	942	951	2,802	0.64
Mathematics	911	912	2,761	0.54
	911	913	2,774	0.54
	911	914	2,791	0.63
	911	915	2,783	0.65
	912	913	2,734	0.52
	912	914	2,756	0.59
	912	915	2,750	0.52
	913	914	2,766	0.53
	913	915	2,764	0.58
	914	915	2,773	0.62
Science	921	931	2,816	0.65
Social Studies	911	921	2,810	0.67