

New York State Alternate Assessment Technical Report 2012–13

Submitted to:

The New York State Education Department



Office of State Assessment

Prepared by:



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

TABLE OF CONTENTS

CHAPTER 1	OVERVIEW	1
1.1	PURPOSE OF THE REPORT	1
1.2	ORGANIZATION OF THIS REPORT.....	2
1.3	CURRENT YEAR UPDATES.....	2
CHAPTER 2	THE STATE ASSESSMENT SYSTEM	3
2.1	INTRODUCTION	3
2.2	ALTERNATE ASSESSMENT BASED UPON ALTERNATE ACHIEVEMENT STANDARDS	4
2.3	THE ALTERNATE ASSESSMENT SYSTEM	5
2.4	PURPOSE OF THE ALTERNATE ASSESSMENT SYSTEM	5
2.5	TEST USE AND DECISIONS BASED ON ALTERNATE ASSESSMENT.....	6
2.6	BACKGROUND AND GENERAL FORMAT	7
2.7	TESTING ACCOMMODATIONS	7
CHAPTER 3	THE STUDENTS	9
3.1	TARGET POPULATION.....	9
3.2	SUMMARY OF PARTICIPATION RATES	10
CHAPTER 4	TEST DEVELOPMENT	13
4.1	FRAMEWORKS OF THE TESTING PROGRAM.....	13
4.2	AGLIS MAPPED TO NYS LEARNING STANDARDS AND CORE CURRICULUM BY GRADE	14
4.3	AGLI SELECTION CRITERIA AND PROCESS.....	15
4.4	TASK DEVELOPMENT.....	17
4.5	AGLI AND TASK REVIEW PROCESS.....	17
CHAPTER 5	TEST CONTENT	18
5.1	ALTERNATE PERFORMANCE LEVEL DESCRIPTORS (APLDs)	18
5.2	ACCESS TO THE GENERAL CURRICULUM	18
5.3	TEST FORMAT.....	19
5.4	ASSESSMENT DIMENSIONS.....	20
CHAPTER 6	ALIGNMENT	22
6.1	DESCRIPTION OF LINKAGES TO DIFFERENT CONTENT ACROSS GRADES.....	22
6.2	PROMOTING ALIGNMENT THROUGH ACHIEVEMENT LEVEL DESCRIPTORS	22
CHAPTER 7	ADMINISTRATION AND TRAINING	25
7.1	STEPS FOR ADMINISTRATION	25
7.2	STEPS IN CONSTRUCTING THE DATAFOLIO	26
7.3	ADMINISTRATION TRAINING AND COLLEGIAL REVIEW	27
CHAPTER 8	SCORING	30
8.1	SCORING OF OPERATIONAL TESTS	30
8.2	SCORING RUBRIC	31
8.3	SCORING PROCESS AND RELIABILITY MONITORING REVIEW.....	34

8.3.1	Scoring Process	34
8.3.2	Reliability Monitoring Review	36
8.4	SCORER QUALIFICATION AND TRAINING	36
8.5	SCORING QUALITY CONTROL	37
CHAPTER 9	CLASSICAL ITEM ANALYSIS	40
9.1	DIFFICULTY AND DISCRIMINATION	40
9.2	STRUCTURAL RELATIONSHIP	42
9.3	BIAS/FAIRNESS.....	43
CHAPTER 10	CHARACTERIZING ERRORS ASSOCIATED WITH TEST SCORES.....	45
10.1	RELIABILITY	45
10.2	SUBGROUP RELIABILITY	47
10.3	DECISION ACCURACY AND CONSISTENCY	47
10.4	INTERRATER CONSISTENCY.....	52
CHAPTER 11	COMPARABILITY (SCALING AND EQUATING)	53
11.1	COMPARABILITY OF SCORES ACROSS YEARS	53
11.1.1	Standard Setting.....	53
11.1.2	Reported Scores (Cumulative Distributions).....	54
11.1.3	Performance Level Distributions	54
11.2	LINKAGES ACROSS GRADES	55
CHAPTER 12	VALIDITY.....	56
12.1	PROCEDURAL VALIDITY	56
12.2	CONTENT VALIDITY	57
12.3	CONSEQUENTIAL VALIDITY	57
REFERENCES		62
APPENDICES		63
APPENDIX A	NYSAA TEST BLUEPRINTS FOR EACH CONTENT AREA	
APPENDIX B	NYSAA SCORING PROCEDURES	
APPENDIX C	2012–13 SCORING DECISION RULES	
APPENDIX D	SUBGROUP RELIABILITY	
APPENDIX E	INTERRATER CONSISTENCY	
APPENDIX F	PERFORMANCE LEVEL DISTRIBUTIONS	
APPENDIX G	CUMULATIVE DISTRIBUTION GRAPHS	

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 assessment’s validity can be evaluated. The Alternate Grade Level Indicators (AGLIs) development process, the alignment of the AGLIs to the New York State learning standards Grade-Level Indicators (GLIs) for English language arts (ELA), mathematics, science, and social studies, which occurred in 2006–07 and 2007–08, are presented in detail. Stakeholder input in the development of the overall NYSAA process itself is described, from the AGLI design through Blueprint/test design, content alignment, assessment task development, teacher trainings, administration, scoring, and standard setting.

1.1 PURPOSE OF THE REPORT

The purpose of this report is to document the technical aspects of the 2012–13 NYSAA. During the 2012–13 school year, approximately 21,571 students in Grades 3 through 8 and in high school participated in the administration of the NYSAA. ELA and mathematics were assessed at the Grades 3 through 8 and high school levels; science was assessed at the Grades 4, 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, APA, & NCME, 1999). Each chapter in this section contributes important information to the validity argument by addressing one or more of the following aspects of the NYSAA: AGLI and assessment task development, alignment, administration, scoring, reliability, standard setting, and achievement levels.

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) provides a framework for describing sources of evidence that should be considered when constructing an argument for assessment validity. These evidence sources include those 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 yearlong process, which includes Blueprint design/development (completed in 2006–07), AGLI development (completed in 2006–07 and 2007–08), assessment task development, administration, scoring, standard setting (completed in 2006–07 and 2007–08), technical characteristics, and validity. The appendices contain supporting documentation.

1.3 CURRENT YEAR UPDATES

The assessment structure and guidelines for the 2012–13 NYSAA remained consistent with previous administrations. There were no changes in assessment requirements, Test Blueprints, or AGLIs for any grade or content area.

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 during high school; on science content taught during Grades 4, 8, and during high school; and on social studies content taught during high school. All students participate in the statewide assessment program through: the general assessments with or without accommodations, the alternate assessment with or without accommodations, or a combination of the general and alternate assessments.

2.1 INTRODUCTION

The New York State Alternate Assessment (NYSAA) is designed to provide a snapshot in time 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 New York State learning 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 Alternate Grade Level Indicators (AGLIs) with the core curriculum in ELA, mathematics, 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 New York State learning standards. 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 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 2012–13 NYSAA was October 1, 2012, through February 8, 2013. The scoring period for the 2012–13 NYSAA was March 11, 2013, to May 2, 2013. The general sequence of events for administering the NYSAA is highlighted 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), provides academic instruction so that the student can achieve proficiency in two different AGLIs in each content area assessed. Two AGLIs are required for each content area assessed (ELA, mathematics, science, and social studies).
3. Parents meet with the teacher to discuss how the NYSAA is administered and which specific AGLIs will be used to assess their child.
4. For each AGLI, the student is required to perform one assessment task connected to the AGLI on three different dates.
5. Members of the student's instructional team conduct assessment tasks and document and rate student performance. This process includes collecting evidence for any two of the three dates of student performance.
6. The teacher assembles a datafolio containing the evidence of student performance and the ratings of the student's Level of Accuracy and Level of Independence. The completed datafolio is submitted to the building administrator on or before the last day of the administration period for shipping to the regional Scoring Institute.
7. The NYSAA datafolios are scored at regional NYSAA Scoring Institutes during the scoring period defined by the Department.
8. Student reports are created and are made available to school districts, teachers, and parents.

2.2 ALTERNATE ASSESSMENT BASED UPON 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 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 New York State Grade-Level Indicators in ELA, mathematics, 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 No Child Left Behind (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 while maintaining linkage 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. In response to a 2005–06 review of the New York State Testing Program by the United States Education Department, the NYSAA was restructured in 2006–07. The 2012–13 administration was the sixth full year of implementation under the redesigned assessment program.

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, the Department aligned the AGLIs (discussed in the following section) with the core curriculum’s Grade-Level Expectations in ELA, mathematics, science, and social studies for the NYSAA administration.

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. Previously, 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 2012–13, 70% agreed that the AGLIs assessed in the NYSAA made the grade-level core curricula more accessible and said 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 2012–13 NYSAA Administration Manual (accessible at <http://www.p12.nysed.gov/assessment/nysaa/archive-12.html>). The process for datafolio development (see Chapter 7) supports the procedural validity for 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 in 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.

Assessment based on AGLIs is accomplished via datafolios. A datafolio is a collection of evidence of a student’s academic performance that 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 New York State learning standards. The CSE can use the datafolio to understand the student’s achievement relative to the New York State learning 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 summer 2008. 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. APLDs, along with 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 rating the student's Level of Accuracy and Level of Independence as he or she performs an assessment task on three different dates 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 all of the assessment requirements, steps for compiling a datafolio, forms, and the NYSAA Frameworks as an appendix. The NYSAA Frameworks include an introduction, and the NYSAA Test Blueprints outline the curriculum content of the alternate assessment for each grade. The Test Blueprints illustrate for each content area (i.e., ELA, mathematics, science, and social studies) the two major areas of curriculum focus that teachers must assess at each grade, referred to as the Required Components. These two Required Components are further divided into specific curriculum topics called Choice Components. Teachers begin constructing a student's datafolio by selecting two of the four Choice Components within each content area's Required Components.

Teachers must identify one AGLI based on the student's assessed grade level for each of the two selected Choice Components at the student's assessed grade level. Thus, teachers must assess a student on two AGLIs for each content area. For each AGLI, the teacher must collect and document student performance data from an assessment task administered on three separate dates. One piece of verifying evidence must be submitted for any two of the three dates of student performance.

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 in 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 Assessment Program 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 following 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 in 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 2012–13 NYSAA by demographic category for each content area.

Table 3-1. 2012–13 NYSAA: Summary of Participation—English Language Arts

<i>Demographic Group</i>	<i>Number Tested</i>	<i>Percent Participation</i>
All Students	21,507	100.00
Male	14,625	68.00
Female	6,882	32.00
American Indian/Alaskan Native	164	0.76
Black	5,490	25.53
Asian	1,111	5.17
Hispanic	5,531	25.72
White	8,995	41.82
Native Hawaiian/Other Pacific Islander	51	0.24
Multi	165	0.77

Table 3-2. 2012–13 NYSAA: Summary of Participation—Mathematics

<i>Demographic Group</i>	<i>Number Tested</i>	<i>Percent Participation</i>
All Students	21,482	100.00
Male	14,603	67.98
Female	6,879	32.02
American Indian/Alaskan Native	164	0.76
Black	5,486	25.54
Asian	1,113	5.18
Hispanic	5,533	25.76
White	8,971	41.76
Native Hawaiian/Other Pacific Islander	51	0.24
Multi	164	0.76

Table 3-3. 2012–13 NYSAA: Summary of Participation—Science

<i>Demographic Group</i>	<i>Number Tested</i>	<i>Percent Participation</i>
All Students	9,076	100.00
Male	6,096	67.17
Female	2,980	32.83
American Indian/Alaskan Native	71	0.78
Black	2,282	25.14
Asian	490	5.40
Hispanic	2,284	25.17
White	3,871	42.65
Native Hawaiian/Other Pacific Islander	17	0.19
Multi	61	0.67

Table 3-4. 2012–13 NYSAA: Summary of Participation—Social Studies

<i>Demographic Group</i>	<i>Number Tested</i>	<i>Percent Participation</i>
All Students	2,970	100.00
Male	1,920	64.65
Female	1,050	35.35
American Indian/Alaskan Native	21	0.71
Black	752	25.32
Asian	172	5.79
Hispanic	685	23.06
White	1,323	44.55
Native Hawaiian/Other Pacific Islander	2	0.07
Multi	15	0.51

CHAPTER 4 TEST DEVELOPMENT

4.1 FRAMEWORKS OF THE TESTING PROGRAM

The New York State learning standards provide the framework for the New York State Testing Program. The grade-level core curricula expand the priorities of the New York State learning standards into grade-level expectations. Each statewide assessment program has a Test Blueprint that outlines the priorities to be assessed based on the grade-level core curricula. The redesign carried out in response to the United States Education Department's 2005–2006 Review of the New York State Testing Program (discussed in Chapter 2) required that the New York State Alternate Assessment (NYSAA) be aligned to grade-level core curricula. 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 fall 2006, the New York State Education Department (the Department) assembled Stakeholders to review the core curriculum and general education assessment Blueprints for English language arts (ELA), mathematics, 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 Stakeholders' 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 Stakeholders expanded the core curriculum Grade-Level Expectations to Alternate Grade Level Indicators (AGLIs) for students with severe cognitive disabilities. The AGLIs now provide an entry point to the grade-level content of the core curriculum so that a student's level can be gauged in terms of the core curriculum established for all students by the New York State Board of Regents.

The Test Blueprints, Grade-Level Expectations, Essences, AGLIs, and Sample Assessment Tasks (SATs) for each grade can be found in the 2012–13 NYSAA Administration Manual: Appendix G—NYSAA Frameworks(www.p12.nysed.gov/assessment/nysaa/archive-12.html).

4.2 AGLIs MAPPED TO NYS LEARNING STANDARDS AND CORE CURRICULUM BY GRADE

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

Stakeholder meetings were held during the summer and early fall of 2006 in order to gather input on aligning the NYSAA requirements with grade-level expectations and on developing AGLIs. Additionally, Stakeholder meetings were held in spring 2007 and 2008 to further refine the AGLIs and to develop additional SATs for teachers to use in the alternate assessment.

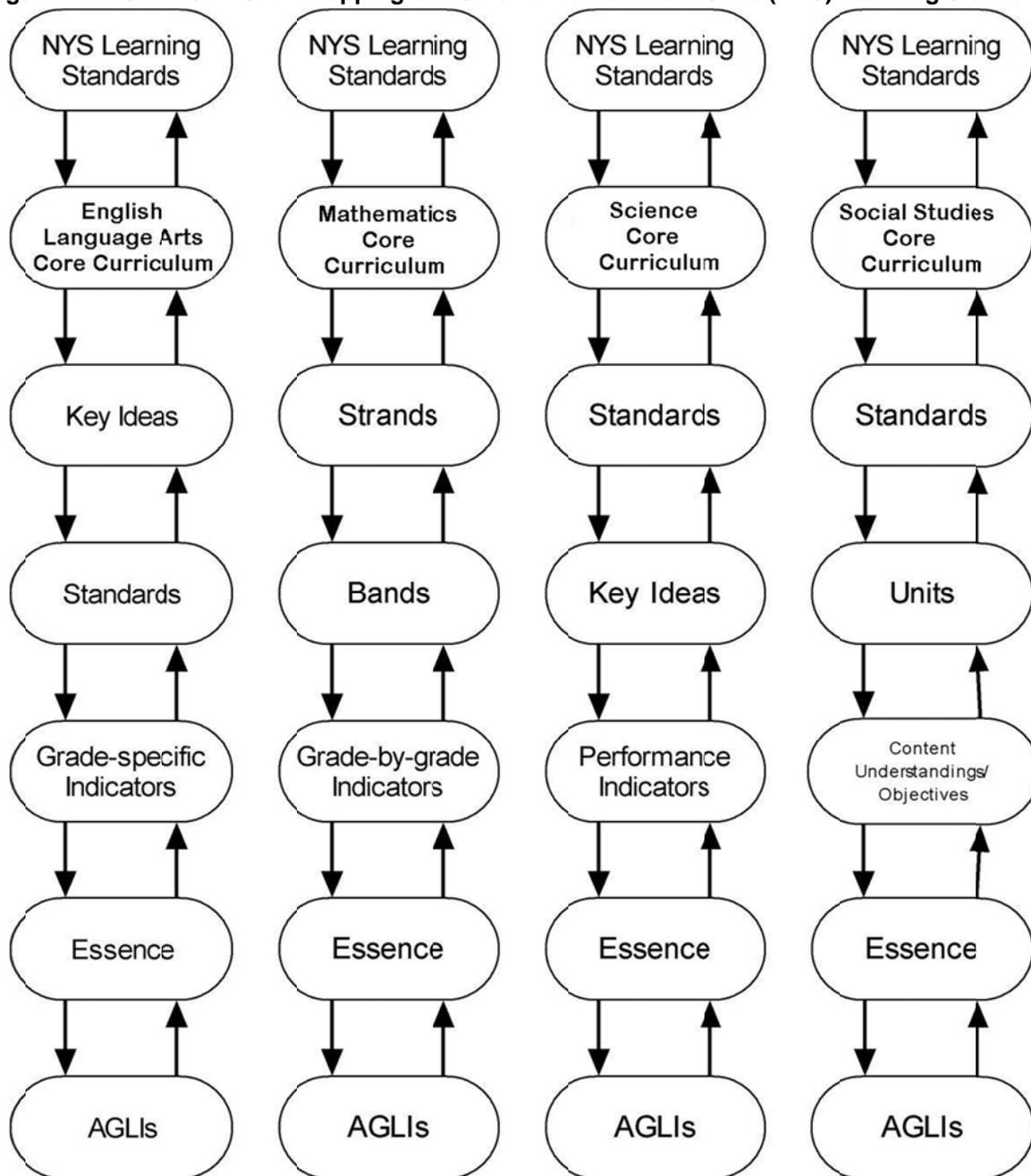
The Board of Regents approved a set of learning standards to guide instruction and assessment. The learning standards serve as the basis of the core curricula in ELA, mathematics, science, and social studies. The curriculum of each content area is divided into the following components:

- English language arts: key ideas and standards
- mathematics: strands and bands
- 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 different levels of complexity.

Figure 4-1. 2012–13 NYSAA: Mapping of AGLIs to the New York State (NYS) Learning Standards



4.3 AGLI SELECTION CRITERIA AND PROCESS

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 the NRWG meetings, which ensured consistency in the overall process and content interpretation.

The NRWG did not meet in spring 2012. There were no edits made to the Test Blueprints, Grade-Level Expectations, Essences, and intent of the AGLIs that were finalized and used in the 2007–08 administration. The 2012–13 NYSAA Frameworks had some updates to SATs and content

glossaries in order to provide clarification and additional information to teachers. However, as was the case with the 2011–12 version of the NYSAA Frameworks, the intent of the AGLIs was not changed in any way.

The spring 2008 NRWG process was consistent across each of the four content areas. 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 work group.

The group 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 they would be working with and explaining the specific tasks for the content area work groups, as well as other logistical information. The group was given time for questions and then released into their content area work groups, 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. 2012–13 NYSAA: NRWG Participant Groups from 2008

<i>Content Area</i>	<i>Group</i>	<i>Grades</i>
English Language Arts	1	3, 4, 5
	2	6, 7
	3	8, High School
Mathematics	1	3, 4, 5, 6
	2	7, 8, High School
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 groups, the participants reviewed and edited existing SATs and then worked to add new SATs. The process for adding new SATs was as follows: The groups first focused on AGLIs that did not have an SAT. Then they developed additional SATs for AGLIs that already had at

least one SAT. Throughout the editing and developing of SATs, each group worked to ensure alignment to the AGLIs. During the editing process, the groups also identified words they felt should be added to the glossary for each content area. The work tasks within each content area focused on each of the identified outcomes for the revision of the NYSAA Frameworks.

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

Each facilitator gathered each group'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 TASK DEVELOPMENT

As part of the redesign process, assessment tasks for the AGLIs were developed, edited, and refined. 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. 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.

4.5 AGLI AND TASK REVIEW PROCESS

The RLTs and Measured Progress reviewed and updated SATs from the 2011–12 NYSAA Frameworks. Revisions were made to existing tasks to clarify their alignments to the AGLIs. New tasks were developed to provide additional samples from which teachers could choose. The Department provided a final content review and approval of the SATs. The AGLIs from the 2011–12 NYSAA Frameworks were not modified in any way. The final AGLIs and SATs can be found in the NYSAA Administration Manual: Appendix G—NYSAA Frameworks (www.p12.nysed.gov/assessment/nysaa/archive-12.html).

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 New York State learning standards are accessible to all students. The assessment design allows students to demonstrate their knowledge, skills, and understanding of the New York State learning standards through the Alternate Grade Level Indicators (AGLIs). The AGLIs are organized into three levels of complexity in order to provide an appropriate entry point for students into the core curricula and maintain the connection with the academic focus of the alternate assessment. Student performance data—Level of Accuracy and Level of Independence—is collected by the teacher for each AGLI that the student is assessed against.

5.1 ALTERNATE PERFORMANCE LEVEL DESCRIPTORS (APLDs)

The Alternate Performance Level Descriptors (APLDs), previously developed during standard setting, were used for the 2012–13 administration and reporting. Standard setting was conducted in June 2008 to establish cut scores for each alternate performance level in English language arts (ELA) and mathematics, Grades 3 through 8 and high school; in science, Grades 4, 8, and high school; and in social studies, Grades 5, 8, and high school.

The June 2007 standard-setting process developed the original APLDs, which were used by the standard-setting groups in June 2008. The APLDs provided panelists with an idea of the knowledge, skills, and understanding related to the core curriculum 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 core curricula 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.

The Department, in cooperation with Stakeholders from across the State, has expanded the core curriculum Grade-Level Performance Expectations to AGLIs for students with severe cognitive

disabilities. AGLIs provide an entry point to the grade-level content of the core curriculum. AGLIs measure a level of mastery of the knowledge, skills, and understanding aligned with the 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. Two components are required for each content area within a grade. Within the Required Components, there are two choices. The Choice Components give the teacher flexibility to assess the student based on specific academic content that was part of the student’s instructional program. This flexibility allows 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.

Tables 5-1 and 5-2 show examples of the Required and Choice Components from the Test Blueprint for ELA contained in the NYSAA Frameworks.

Table 5-1. 2012–13 NYSAA: ELA Required Components (Two per Grade Level)

<i>English Language Arts Key Idea</i>	<i>Grade 3</i>	<i>Grade 4</i>	<i>Grade 5</i>	<i>Grade 6</i>	<i>Grade 7</i>	<i>Grade 8</i>	<i>High School</i>
Reading	X	X	X	X	X	X	X
Writing		X		X		X	X
Listening Speaking*	X		X		X		

* Speaking is not assessed on the general education state assessments.

**Table 5-2. 2012–13 NYSAA: ELA Choice Components
(One Standard Chosen per Each Required Key Idea per Grade Level)**

<i>Key Idea</i>	<i>Standard</i>	<i>Grade 3</i>	<i>Grade 4</i>	<i>Grade 5</i>	<i>Grade 6</i>	<i>Grade 7</i>	<i>Grade 8</i>
Reading	1			X	X	X	X
	2	X	X	X	X	X	
	3						X
	4	X	X				
Writing	1		X		X		X
	2		X		X		
	3						X
	4						
Listening	1			X		X	
	2	X		X		X	
	3						
	4	X					

A datafolio is the resulting body of evidence across Required and Choice Components of a student's academic performance of selected AGLIs, as compiled by the student's instructional team and scored by qualified Scorers. The teacher is required to select one AGLI from each Required Component by which to assess the student. Teachers are not allowed to modify AGLIs. Student performance is rated by the student's instructional team according to the student's Levels of Accuracy and Independence in performing each assessment task. This is done on three separate dates within the administration period. Teachers have three options in determining the assessment tasks: (1) use a Sample Assessment Task (SAT) from the NYSAA Frameworks, (2) modify an SAT from the NYSAA Frameworks to make it more applicable to the student's specific needs, abilities, and/or mode of communication, or (3) create an original assessment task. The assessment task is the student action and is aligned to the AGLI the student is being assessed against. To verify this documentation, each datafolio must include the following: student work products, Data Collection Sheets, photographs, and/or digital video or audio recordings for two of the three dates of documented 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 are 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 assessment task is clearly aligned with the AGLI and
- the verifying evidence submitted is aligned with the assessment task.

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

- **Performance**

- Level of Accuracy
- Level of Independence

Both the Level of Accuracy and Level of Independence are components of the performance dimension and are calculated as a percentage (0%–100%) and then rated on a scale of 1, 2, 3, or 4 (see Table 5-3).

Table 5-3. 2012–13 NYSAA: Scoring Rubric

<i>Level</i>	<i>Rating</i>
100%–80%	4
79%–60%	3
59%–30%	2
29%–0%	1

CHAPTER 6 ALIGNMENT

6.1 DESCRIPTION OF LINKAGES TO DIFFERENT CONTENT ACROSS GRADES

The independent alignment study conducted in April 2007 of the New York State Alternate Assessment (NYSAA) for the New York State Education Department (the Department) provided information about the alignment of the New York State learning standards found in the core curricula, the Alternate Grade Level Indicators (AGLIs), and the Sample Assessment Tasks (SATs). The alignment study found that AGLI content is academic and addresses the major domains and strands of the core curricula as compared to State and national standards in English language arts (ELA), mathematics, and science. It was also found that the distribution of AGLIs across ELA, mathematics, and science content is consistent at each grade level to the NYSAA Test Blueprints. The alignment study used a modification of the alignment system proposed by Flowers, Browder, Wakeman, and Karvonen (2006) to form the basis of the study. Webb's depth-of-knowledge (DOK) classification system (2002) and the refinements to this system suggested by Tindal (2005) to address the unique features of an alternate assessment, as incorporated into the Flowers et al. design, was a focal point of the alignment study. The DOK levels for AGLIs and SATs reflected the DOK levels in the core curriculum. Lastly, the alignment study found that in the scored datafolios reviewed, there was a strong relationship between the AGLIs and the assessment tasks, as well as overall compliance with the NYSAA administration guidelines.

6.2 PROMOTING ALIGNMENT THROUGH ACHIEVEMENT LEVEL DESCRIPTORS

The Alternate Performance Level Descriptors (APLDs) for the NYSAA are uniquely defined for each grade and content area with unifying adverbs. The APLDs provide a structure for understanding the knowledge, skills, and understanding that a student may have demonstrated in the 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 APLDs development occurred in 2007 and 2008 as part of the NYSAA redesign. It began with reviewing and utilizing text from the general education performance level descriptors. The general education performance level descriptors used terms such as *limited*, *basic*, *proficiency*, and

consistently. The initial draft of the APLDs used similar terms and was brought to the NYSAA Advisory Committee members for further refinement based on their in-depth understanding and experience with students with severe cognitive disabilities and the knowledge, skills, and understanding the students demonstrate in the datafolios. The edits that the Advisory Committee proposed utilized definitions that provided a picture of a student's accuracy performance and independence performance as related to the AGLIs as a whole within a content area. The recommended terms for the performance level of Not Meeting Learning Standards were *limited* and *rarely*. The recommended terms for the performance level of Partially Meeting Learning Standards were *basic* and *inconsistent*. The recommended terms for the performance level of Meeting Learning Standards were *often* and *basic*. The recommended terms for the performance level of Meeting Learning Standards with Distinction were *consistently* and *frequently*. These terms were used to develop the final drafts of the APLDs that were utilized during standard setting by each of the groups. The groups were asked to further refine the APLDs, keeping in mind the variations that students with severe cognitive disabilities may demonstrate in their work as evidenced in a datafolio. The groups used the draft APLDs, in coordination with student datafolio samples, to come to a consensus on the words and information that were to be included in the final APLDs. Each of the groups recommended that consistent wording be included throughout each of the APLDs for each grade and content area. The terms *rarely*, *inconsistently*, *often*, and *consistently* were the final recommendations from the standard-setting panelists. Rigid definitions of these terms are not particularly useful in understanding each of the performance levels, especially for this population. It is important to understand the terms and to review them within the context of the APLDs.

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 core curricula. 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 and independence levels. Many students who take the NYSAA have splinter skills, require various supports in order 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.

The APLDs provided the standard-setting panelists the official description of the knowledge, skills, and understandings students are expected to be able to display to be classified into each performance level. Panelists were given an opportunity at the end of the standard-setting process to recommend additions and refinements to the APLDs.

The APLDs were finalized and approved by the Department following the standard setting that took place in June 2008. The APLDs have remained unchanged since then.

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 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 need to also have an understanding of the NYSAA administration guidelines.

7.1 STEPS FOR ADMINISTRATION

The teacher, in coordination with the instructional team, is responsible for the administration of the NYSAA with 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 in which the assessment can be conducted. Assessment data cannot be collected before or after the administration period. The administration period for 2012–13 was October 1, 2012, to February 8, 2013. 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 the student will be assessed at, using the student's date of birth and the NYSAA Age Range Chart. Next, the teacher determines the Alternate Grade Level Indicator (AGLI) for each Required Component that the student will be assessed against. There are two Required Components within each content area. Then, the teacher determines an assessment task that will demonstrate the AGLI. The assessment task describes the student action being assessed. Once the AGLIs and assessment tasks have been determined, the teacher conducts the assessment task with the student on at least three different dates. The last three dates of performance of the AGLI and assessment task are recorded and documented. Student performance includes the student's Level of Accuracy and Level of Independence. Verifying evidence showing the student demonstrating the knowledge, skills, and understanding of the AGLI through the completion of the assessment task must be included for any two of the three 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 AGLI, there must be a Data Summary Sheet. The Data Summary Sheet is the summarizing information regarding the AGLI. It includes student demographic information, the AGLI assessed, the assessment task, and student performance data. The three dates of performance data include both the percentages and corresponding Scoring Rubric rating for both the Level of Accuracy and the Level of Independence. In addition to the individual requirements of each type of verifying evidence, the verifying evidence must include seven required elements—student name, date of student performance, content area, AGLI text, assessment task, Level of Accuracy, and Level of Independence. 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, as well as the grade assessed, supports required per the IEP, testing accommodations provided during testing, and the month a Collegial Review was conducted. A datafolio also includes a table of contents, which provides information to Scorers on where information is located in the datafolio. The English language arts (ELA) assessment documents come first, followed by mathematics, then science and social studies, if applicable. The AGLI documentation is organized by Required Components. It is the same for each content area.

First Required Component—AGLI 1:

- Data Summary Sheet for this AGLI and
- two pieces of verifying evidence

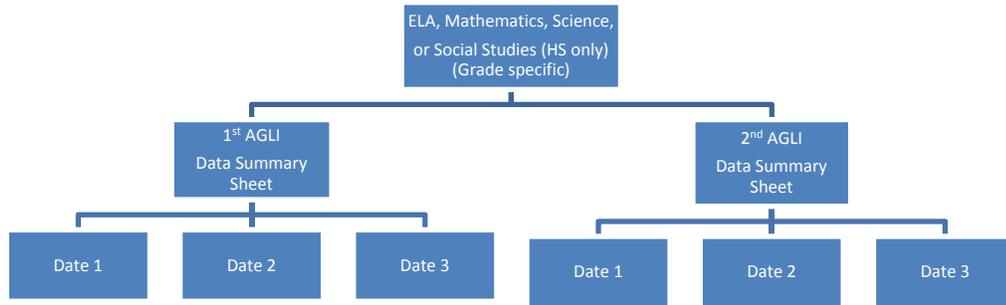
If either piece of verifying evidence is a Data Collection Sheet, supporting evidence is placed directly behind the Data Collection Sheet.

Second Required Component—AGLI 2:

- Data Summary Sheet for this AGLI and
- two pieces of verifying evidence

If either piece of verifying evidence is a Data Collection Sheet, supporting evidence is placed directly behind the Data Collection Sheet.

Figure 7-1. 2012–13 NYSAA: Datafolio Elements



7.3 ADMINISTRATION TRAINING AND COLLEGIAL REVIEW

In September 2012, the New York State Education Department (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 were provided. Then, portions of the NYSAA Administration Training DVD were shown, as well as completion and review of the Guided Practices. Last, the participants were asked to work in groups to discuss best practice tips and strategies for administering the NYSAA.

A total of five NYSAA Administration Updates 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. 2012–13 NYSAA: Administration Updates Training—Participant Count

	<i>Albany Region</i>	<i>Syracuse Region</i>	<i>Buffalo- Rochester Region</i>	<i>New York City Region (Two Trainings)</i>	<i>Total</i>
NYSAA Administration Updates Training	54	22	45	116	237

Administration Training to 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 new teacher segment, a segment devoted to updates and a review for all teachers, and a closing segment. The opening segment provides general information about what is going to be covered during the training session. The new teacher segment is a detailed review of each of the steps for administering the NYSAA. The updates and review for All Teachers' Segment provides updates regarding the assessment, a scored datafolio review, information on how to maintain the Connection to Grade-Level Content during administration, information on Data Collection Sheets, and reminders regarding the NYSAA. The closing segment outlines information regarding next steps for teachers and provides information regarding Collegial Reviews. The DVD is set up to allow AATN Specialists to provide a training session for new teachers only, for experienced teachers only, or for a combination of new and experienced teachers. At specific points throughout the segments, there are stop points built in, and a Guided Practice must be conducted at each of these points. 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 determine the correct grade a student should be assessed at based on his or her date of birth, and how to use the NYSAA Administration Manual Appendix G: NYSAA Frameworks (www.p12.nysed.gov/assessment/nysaa/archieve-12.html), to navigate through the NYSAA Frameworks in order to select the AGLIs. The second Guided Practice provides teachers an opportunity to complete Data Summary Sheets and see how a content area entry would be organized in a datafolio. It also includes a general NYSAA requirements review worksheet. The third Guided Practice focuses on determining and documenting student performance data. The fourth Guided Practice provides teachers with practice in understanding the Connection to Grade-Level Content. Primarily, it provides examples of verifying evidence that align to the assessment task and examples of verifying evidence that do not align to the assessment task. Teachers who are new complete all four Guided Practices. Teachers who are experienced complete only Guided Practices three and four. At or before the locally conducted NYSAA Administration Trainings, teachers are provided with the NYSAA Administration Manual, which includes the NYSAA Frameworks as Appendix G.

Collegial Review is required for each student datafolio. Collegial Review is meant to be an independent review of a datafolio. The Department provides guidance that reviewers should:

- be familiar with the current alternate assessment;
- have attended training in fall 2012; and/or
- be experienced and qualified as a Scorer in spring 2012 (the 2011–12 NYSAA scoring).

The Department guidelines recommend 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 AGLIs and the

assessment tasks and between the assessment tasks and the verifying evidence. Also, Collegial Reviews 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 in order 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 the Scoring Institute. The AATN Specialist acts as a Floor Manager at a Scoring Institute and must also pass the qualification samples in order to make content decisions during the Scoring Institute.

In February and March 2013, the New York State Education Department (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. There were three main activities conducted. First, information regarding updates to the NYSAA and the materials were provided. Then, the NYSAA Scoring Training DVD was shown, as well as completion and review of the practice samples. Last, the participants were asked to complete the qualification samples. Retraining and qualification was provided as needed.

A total of five NYSAA Scoring Trainings occurred at four geographically diverse sites: the Albany region, which includes the Long Island and surrounding New York City regions; the Syracuse region; the Buffalo and Rochester region; and the New York City region, which includes the non-District 75 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. 2012–13 NYSAA: Scoring Training—Participant Count

	<i>Albany Region</i>	<i>Syracuse Region</i>	<i>Buffalo- Rochester Region</i>	<i>New York City Region (Two Trainings)</i>	<i>Totals</i>
NYSAA Scoring Training	43	27	48	102	220

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 in which educators individually 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 is satisfied, and verify the percentages and ratings for Accuracy and Independence documented by the teacher for each Alternate Grade-Level Indicator (AGLI) assessed. Any questions that arise during scoring are directed to a Table Leader. Scorers use the document entitled *Steps for Scoring 2012–13 NYSAA Datafolios* as the main reference for scoring each datafolio. Table Leaders use the *Decision Rules for Scoring 2012–13 NYSAA Datafolios* as a reference document for any questions that are not addressed in the *Steps for Scoring 2012–13 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 AGLI code, Connection to Grade-Level Content questions, ratings for the Levels of Accuracy and Independence, and 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 AGLI codes, Connection to Grade-Level Content questions, ratings, 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 2012–13 NYSAA Administration Manual, along with guidance on the process that teachers must follow in order to meet the scoring requirements. The rubric is broken into two parts. The first part outlines the content and grade-level Required Components. The second part is the rating summary. The rating is based on the Connection to Grade-Level Content and student performance. The Connection to Grade-Level Content is explained on the Scoring Rubric as follows: “AGLIs are the expansion of the academic content for students with severe cognitive disabilities. The assessment task must connect to the AGLI and the verifying evidence must demonstrate the task. If these connections are not clear, the AGLI will not be scored.” For each assessment task documented, the ratings for Level of Accuracy and Level of Independence (relative to the student’s demonstration of

skills, in relation to the AGLI) combine to give the performance dimension. The Scoring Rubric is presented in Table 8-2.

Table 8-2. 2012–13 NYSAA: Scoring Rubric

For each content area at each grade, two AGLIs must be assessed on three dates within the administration period.

Charted below are the two Required Components for each grade and content area.

<i>Content Area</i>	<i>Grade 3</i>	<i>Grade 4</i>	<i>Grade 5</i>	<i>Grade 6</i>	<i>Grade 7</i>	<i>Grade 8</i>	<i>High School</i>
English Language Arts	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Listening 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Writing 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Listening 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Writing 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Listening 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Writing 	<ul style="list-style-type: none"> • Key Idea Reading • Key Idea Writing
Mathematics	<ul style="list-style-type: none"> • Strand Number Sense and Operations • Strand Measurement 	<ul style="list-style-type: none"> • Strand Number Sense and Operations • Strand Measurement 	<ul style="list-style-type: none"> • Strand Number Sense and Operations • Strand Geometry 	<ul style="list-style-type: none"> • Strand Number Sense and Operations • Strand Algebra 	<ul style="list-style-type: none"> • Strand Number Sense and Operations • Strand Statistics and Probability 	<ul style="list-style-type: none"> • Strand Geometry • Strand Algebra 	<ul style="list-style-type: none"> • Strand Algebra • Strand Statistics and Probability
Science		<ul style="list-style-type: none"> • Standard 1 Scientific Inquiry • Standard 4 Living Environment or Physical Setting/ Earth Science 				<ul style="list-style-type: none"> • Standard 1 Scientific Inquiry • Standard 4 Living Environment or Physical Setting/ Earth Science 	<ul style="list-style-type: none"> • Standard 4 Living Environment • Standard 4 Physical Setting/ Earth Science
Social Studies							<ul style="list-style-type: none"> • Standard 1 U.S. History • Standard 2 Global History

CONNECTION TO GRADE-LEVEL CONTENT + PERFORMANCE = RATING

Connection to Grade-Level Content – AGLIs are the expansion of the academic content for students with severe cognitive disabilities.

The assessment task must connect to the AGLI and the verifying evidence must demonstrate the task. If these connections are not clear, the AGLI will not be scored.

Performance = Level of Accuracy + Level of Independence

<i>RATING</i>	<i>4</i>	<i>3</i>	<i>2</i>	<i>1</i>	<i>No Score (NS)</i>
Level of Accuracy	The student demonstrates skills based on AGLIs with an average of 80–100% Accuracy.	The student demonstrates skills based on AGLIs with an average of 60–79% Accuracy.	The student demonstrates skills based on AGLIs with an average of 30–59% Accuracy.	The student demonstrates skills based on AGLIs with an average of 0–29% Accuracy.	Required evidence of student performance was not submitted. OR Scorer was unable to determine a score based on the submitted evidence.
Level of Independence	The student seldom requires cues or prompts when demonstrating skills based on the documented AGLIs.(80–100% Independence)	The student requires limited cues or prompts to demonstrate skills based on the documented AGLIs.(60–79% Independence)	The student requires extensive cues or prompts to demonstrate skills based on the documented AGLIs.(30–59% Independence)	The student requires constant cues or prompts to demonstrate skills based on the documented AGLIs.(0–29% Independence)	Required evidence of student performance was not submitted. OR Scorer was unable to determine a score based on the submitted evidence.

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 practice samples, and Scorer qualification. (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 requirements are met. The Scorer records the rubric rating for each AGLI assessed. If the Connection to Grade-Level Content is satisfied, the performance percentages can be confirmed, and each performance percentage for each date is given a rating of 4, 3, 2, or 1. If the Connection to Grade-Level Content is not met, 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, Scoring Procedures, and Scoring Decision Rules. 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 AGLI has a No Score rating. It also allows the Scorer to provide additional constructive feedback to a teacher about the datafolio.

There are 13 steps involved in the scoring process. The step-by-step procedures outlined in the Steps for Scoring 2012–13 NYSAA Datafolios are implemented statewide and ensure scoring reliability across all Scoring Institutes. Table 8-3 presents a quick review of the steps.

Table 8-3. 2012–13 NYSAA: Scoring Steps Quick Reference

<i>Step(s)</i>	
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
3a and b	Demographic and Choice Component information complete on DSS
4a and b	AGLI from grade level (Connection to Grade-Level Content)
5a and b	Task connects to AGLI (Connection to Grade-Level Content)
6a, b, and c	Verifying evidence connects to task (Connection to Grade-Level Content)
7a, b, and c	Dates on DSS: Three separate dates, in chronological order, correspond to dates on verifying evidence
8a, b, and c	Valid verifying evidence and supporting evidence
a	Valid verifying evidence and supporting evidence: Required elements clearly documented (7)
b	Valid verifying evidence and supporting 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 and supporting evidence: Photographs: Minimum of three sequential, captioned, and dated photographs
f	Valid verifying evidence and supporting evidence: digital video or audio clip: Maximum 90 seconds and recorded markers
9	Confirm ratings for Levels of Accuracy and Independence
10	Score the second AGLI (Steps 3–9)
11	Score mathematics, science, and social studies (Steps 2–10)
12	Confirm Scorer Worksheet is complete, including Procedural Error Comments and additional Scorer Comments
13	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 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 2012–13 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 Table Leader segment of 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 the direction provided in the Scoring Procedures document. The rules are organized by topic, beginning with “Incorrect or teacher-created NYSAA forms were used,” “Assessment Task,” “Verifying Evidence,” and “Dates.” Fifteen Decision Rules were developed 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, highlighting 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 2012–13 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 2012–13, the scoring period was March 11–May 2, 2013. 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 introduction, which briefly outlines 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.

After the first two DVD segments, Scorers practice scoring—first as a group, then in pairs, and then individually. Each practice is reviewed to ensure that Scorers are following the Scoring Procedures accurately. The final DVD segment details the subsequent steps in Scorer Training and explains how student scores are reported.

After the DVD, Scorers participate in an activity that reinforces what they have learned about the Scoring Procedures. Then, they are given an opportunity for final questions. Training ends with Scorers completing three calibrated qualifiers. The qualifiers are actual student datafolios in a content area. The qualifiers were identified by a group of Stakeholders 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

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 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 and managing 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 security of all datafolios during transit, storage, and scoring;
- 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;
- 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; 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 2012–13 NYSAA;
- complete Scoring Training, including the qualification process, prior to the start of the Scoring Institute;
- manage scoring at their assigned scoring stations;
- resolve Scorer questions by using scoring guidelines and resources;
- review all corrections and all NS ratings 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 box; and
- provide feedback to SSCs and the Department about the scoring processes, procedures, and documentation.

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. The second Scorer scores the datafolio, completes a second Scorer Worksheet, and returns the datafolio to the original Table Leader. The Table Leader compares the two worksheets. If no discrepancy exists, the Table Leader at the first scoring station fills in his or her Scorer ID# and completes the Scannable. A quality control check is completed, a blue dot is affixed to the datafolio, and the datafolio is returned to the box. 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 for resolution. The Floor Manager reviews the discrepant areas, enters his or her Scorer ID#, and completes the Scannable. The Floor Manager returns the datafolio to the Table Leader at the first scoring station. 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 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 the processes and guidelines prescribed by the Department, a score site observation visit is conducted on a sample of Scoring Institutes. 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) assigned to the particular region. 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 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, APA, NCME, 1999) 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—assessment, 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 2012–13 NYSA.

9.1 DIFFICULTY AND DISCRIMINATION

For the NYSA, each student datafolio for a specified content area at a given grade level receives an Accuracy score and an Independence score. Each of these measurements is taken at three points within the administration period, which results in 12 subscores that are summed to yield a student’s total score, referred to here as a test score. These subscores can be considered similar to traditional test items. Using this definition, 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. Students can achieve a score of 1, 2, 3, or 4 for both Accuracy and Independence. 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 very low difficulty index are considered to be potentially problematic because they are either so difficult that few students get them right or so easy that nearly all students get them right. In either case, such items should be reviewed for appropriateness for inclusion on the assessment. If an assessment were

composed entirely of very easy or very hard items, all students would receive nearly the same scores, and the assessment would not be able to differentiate high-ability students from low-ability students.

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 the 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 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 Table 9-1. The mean difficulty values shown in the table indicate that, overall, students performed well on the items on the NYSAA. 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 Table 9-1 should not be interpreted to mean that the students performed better on the NYSAA than the students who took general assessments did on those tests.

Also shown in Table 9-1 are the mean discrimination values. Because the majority of students received high scores on the tasks, the discrimination indices are somewhat lower than one might expect. This is an artifact of how discrimination is calculated: if all of the students get an item correct, there is little variability in the criterion scores that can be differentiated. 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 Table 9-1 should be interpreted with caution.

Table 9-1. 2012–13 NYSAA: Summary of Item Difficulty and Discrimination Statistics by Content Area and Grade

Content Area	Grade	Number of Items	p-Value		Discrimination	
			Mean	Standard Deviation	Mean	Standard Deviation
English Language Arts	3	12	0.96	0.01	0.49	0.08
	4	12	0.96	0.01	0.47	0.09
	5	12	0.97	0.01	0.48	0.12
	6	12	0.96	0.01	0.52	0.11
	7	12	0.96	0.01	0.48	0.11
	8	12	0.95	0.01	0.53	0.11
	High School	12	0.94	0.01	0.49	0.13
Mathematics	3	12	0.95	0.01	0.57	0.11
	4	12	0.95	0.01	0.56	0.09
	5	12	0.96	0.01	0.56	0.13
	6	12	0.95	0.01	0.58	0.08
	7	12	0.96	0.01	0.52	0.11
	8	12	0.95	0.01	0.52	0.09
	High School	12	0.95	0.02	0.51	0.13
Science	4	12	0.97	0.01	0.44	0.10
	8	12	0.96	0.01	0.45	0.11
	High School	12	0.95	0.01	0.50	0.12
Social Studies	High School	12	0.95	0.01	0.52	0.10

9.2 STRUCTURAL RELATIONSHIP

By design, the performance level classification of the NYSAA is based on two dimensions (Accuracy and Independence). As with any assessment, it is important that these dimensions be carefully examined. This was achieved by exploring the relationships between student dimension scores 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.

The average correlations between Accuracy and Level of Independence by content area and grade are shown in Table 9-2.

Table 9-2. 2012–13 NYSAA: Average Correlations

<i>Content Area</i>	<i>Grade</i>	<i>Number of Items</i>	<i>Average Correlation</i>	<i>Correlation Standard Deviation</i>
English Language Arts	3	6	0.25	0.02
	4	6	0.19	0.07
	5	6	0.20	0.02
	6	6	0.17	0.03
	7	6	0.18	0.03
	8	6	0.13	0.03
	High School	6	0.12	0.03
Mathematics	3	6	0.14	0.04
	4	6	0.14	0.03
	5	6	0.14	0.02
	6	6	0.20	0.02
	7	6	0.15	0.02
	8	6	0.15	0.04
	High School	6	0.13	0.05
Science	4	6	0.16	0.02
	8	6	0.13	0.02
	High School	6	0.13	0.03
Social Studies	High School	6	0.13	0.02

The average correlations between the two dimensions are low to moderate. In this case, a low positive correlation (near zero) indicates that students' Level of Independence had little relationship with their Level of Accuracy, which is a desirable result. A high positive correlation would indicate that students who were able to complete the tasks independently were better able to complete the task (i.e., they had higher Accuracy scores) than students who required more assistance. However, the purpose of providing assistance to students is to enable them to be able to better demonstrate what they know and are able to do. The low to moderate correlations suggest that the assistance provided to students is having the desired effect.

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 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, APA, NCME, 1999).

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 for all students. Chapter 8 describes in detail the Scoring Rubrics used, selection and training of Scorers, and scoring quality control procedures. These processes were followed in order 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 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, scores are defined in the same way as described in Chapter 9. Specifically, the 12 subscores that are awarded for each datafolio (three Accuracy scores and three Independence scores) are treated as the item scores.

10.1 RELIABILITY

In the previous chapter, individual item characteristics of the 2012–13 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, subscore units) 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 on them 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 2012–13 NYSA A tests. The formula is as follows:

$$\alpha \equiv \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_{(Y_i)}^2}{\sigma_x^2} \right]$$

where

i indexes the item,

n is the number of items,

$\sigma_{(Y_i)}^2$ represents individual item variance, and

σ_x^2 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 subscore units complement one another and suggests that the assessment is reliable. Table 10-1 presents raw score descriptive statistics (maximum possible score, average, and standard deviation), Cronbach's α coefficient, and raw score standard errors of measurements (SEMs) for each content area and grade. Because each subscore ranged from 1 to 4, and there were only 12 subscores summed to obtain the total test score, the estimated reliability coefficients were, as expected, somewhat lower than would be found with the typical general assessment, whose reliability coefficients tend to be near 0.90. Considering that the NYSA A assessments are necessarily shorter than general assessments, the reliability coefficients in Table 10-1 are probably comparable.

Table 10-1. 2012–13 NYSA A: Raw Score Descriptive Statistics, Cronbach's Alpha, and Standard Errors of Measurement (SEM) by Grade and Content Area

Content Area	Grade	Number of Students	Raw Score			Reliability (α)	SEM
			Maximum	Mean	Standard Deviation		
English Language Arts	3	2,748	48	45.24	5.84	0.86	2.22
	4	3,019	48	45.26	5.81	0.84	2.34
	5	3,211	48	45.55	5.32	0.83	2.17
	6	3,295	48	45.25	5.92	0.85	2.30
	7	3,155	48	45.23	6.01	0.83	2.48
	8	3,111	48	45.06	5.81	0.85	2.26
	High School	2,968	48	44.22	6.43	0.84	2.56
Mathematics	3	2,749	48	45.33	5.28	0.86	2.00
	4	3,015	48	45.33	5.38	0.85	2.05
	5	3,202	48	45.71	5.06	0.85	1.94
	6	3,298	48	44.72	6.72	0.88	2.28
	7	3,158	48	45.23	5.79	0.84	2.28
	8	3,111	48	44.81	6.46	0.85	2.51
	High School	2,949	48	44.47	6.49	0.85	2.53
Science	4	3,012	48	45.59	5.66	0.84	2.30
	8	3,100	48	44.61	6.77	0.84	2.69
	High School	2,964	48	44.85	6.02	0.84	2.40

Content Area	Grade	Number of Students	Raw Score			Reliability (α)	SEM
			Maximum	Mean	Standard Deviation		
Social Studies	High School	2,970	48	45.08	5.85	0.84	2.33

10.2 SUBGROUP RELIABILITY

The reliability coefficients discussed in the previous section were based on the overall population of students who took the 2012–13 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 only reported 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 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, it can be readily seen in Appendix D 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). Third, 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 and 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) technique was used for the NYSAA because it is easily adaptable to all types of testing formats.

The accuracy and consistency estimates reported in the tables below 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 and 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 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 0.76 for Not Meeting Learning Standards for Grade 3 English language arts (ELA). This figure indicates that among the students whose true scores placed them in this classification, 76% would be expected to be in this classification when categorized according to their observed scores. Similarly, a consistency value of 0.71 indicates that 71% of students with observed scores in the Not Meeting Learning Standards level would be expected to score in this classification again if a second, parallel test form were used.

Table 10-2. 2012–13 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	0.90 (0.88)	0.57	0.76 (0.71)	0.47 (0.37)	0.40 (0.30)	0.98 (0.96)
	4	0.86 (0.82)	0.48	0.75 (0.71)	0.49 (0.39)	0.36 (0.22)	0.95 (0.92)
	5	0.85 (0.80)	0.45	0.69 (0.61)	0.53 (0.43)	0.50 (0.36)	0.94 (0.91)
	6	0.86 (0.84)	0.51	0.72 (0.62)	0.62 (0.53)	0.37 (0.29)	0.97 (0.93)
	7	0.92 (0.90)	0.54	0.56 (0.41)	0.65 (0.57)	0.38 (0.27)	0.98 (0.96)
	8	0.91 (0.89)	0.56	0.62 (0.44)	0.65 (0.55)	0.52 (0.43)	0.98 (0.96)
	High School	0.88 (0.85)	0.54	0.69 (0.58)	0.50 (0.39)	0.51 (0.43)	0.97 (0.95)
Mathematics	3	0.89 (0.86)	0.61	0.62 (0.43)	0.66 (0.56)	0.64 (0.59)	0.97 (0.94)
	4	0.83 (0.79)	0.50	0.69 (0.52)	0.69 (0.62)	0.54 (0.47)	0.94 (0.89)
	5	0.85 (0.82)	0.51	0.66 (0.50)	0.64 (0.54)	0.58 (0.50)	0.95 (0.91)
	6	0.89 (0.86)	0.59	0.73 (0.62)	0.61 (0.51)	0.60 (0.53)	0.97 (0.94)
	7	0.82 (0.78)	0.45	0.76 (0.69)	0.39 (0.30)	0.50 (0.40)	0.93 (0.89)
	8	0.81 (0.77)	0.47	0.75 (0.68)	0.42 (0.33)	0.49 (0.40)	0.94 (0.89)
	High School	0.86 (0.83)	0.58	0.58 (0.39)	0.68 (0.59)	0.60 (0.55)	0.97 (0.92)
Science	4	0.92 (0.90)	0.55	0.67 (0.59)	0.32 (0.25)	0.54 (0.42)	0.98 (0.96)
	8	0.89 (0.86)	0.54	0.60 (0.53)	0.54 (0.45)	0.35 (0.24)	0.98 (0.96)
	High School	0.89 (0.86)	0.55	0.57 (0.40)	0.61 (0.52)	0.52 (0.42)	0.97 (0.95)
Social Studies	High School	0.87 (0.84)	0.54	0.75 (0.68)	0.41 (0.31)	0.48 (0.40)	0.97 (0.94)

For some testing situations, of 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 tables above 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. Summary of Decision (and Consistency) Results
by Content Area and Grade—Conditional on Cutpoint**

Content Area	Grade	<i>Not Meeting / Partially Meeting</i>			<i>Partially Meeting / Meeting</i>			<i>Meeting / Meeting with Distinction</i>		
		<i>Accuracy (consistency)</i>	<i>False</i>		<i>Accuracy (consistency)</i>	<i>False</i>		<i>Accuracy (consistency)</i>	<i>False</i>	
			<i>Positive</i>	<i>Negative</i>		<i>Positive</i>	<i>Negative</i>		<i>Positive</i>	<i>Negative</i>
English Language Arts	3	0.98 (0.97)	0.01	0.01	0.97 (0.96)	0.02	0.01	0.94 (0.93)	0.04	0.02
	4	0.98 (0.97)	0.02	0.01	0.96 (0.95)	0.02	0.01	0.91 (0.88)	0.05	0.04
	5	0.98 (0.98)	0.01	0.01	0.97 (0.96)	0.02	0.01	0.89 (0.85)	0.06	0.05
	6	0.99 (0.98)	0.01	0.01	0.96 (0.94)	0.03	0.02	0.91 (0.89)	0.06	0.03
	7	0.99 (0.99)	0.00	0.01	0.97 (0.96)	0.02	0.01	0.95 (0.93)	0.03	0.02
	8	0.99 (0.99)	0.00	0.00	0.97 (0.96)	0.02	0.01	0.94 (0.93)	0.04	0.02
	High School	0.98 (0.97)	0.01	0.01	0.96 (0.95)	0.02	0.01	0.93 (0.91)	0.05	0.02
Mathematics	3	0.99 (0.99)	0.00	0.00	0.98 (0.97)	0.01	0.01	0.92 (0.90)	0.06	0.02
	4	0.99 (0.99)	0.00	0.00	0.96 (0.95)	0.02	0.02	0.87 (0.84)	0.09	0.04
	5	0.99 (0.99)	0.00	0.00	0.97 (0.96)	0.01	0.01	0.89 (0.86)	0.07	0.04
	6	0.99 (0.99)	0.00	0.01	0.97 (0.96)	0.02	0.01	0.92 (0.90)	0.05	0.02
	7	0.98 (0.97)	0.01	0.01	0.97 (0.96)	0.02	0.01	0.87 (0.84)	0.08	0.05
	8	0.97 (0.97)	0.01	0.01	0.96 (0.95)	0.03	0.02	0.87 (0.84)	0.09	0.04
	High School	0.99 (0.99)	0.00	0.00	0.97 (0.95)	0.02	0.01	0.90 (0.88)	0.08	0.02
Science	4	0.98 (0.98)	0.01	0.01	0.98 (0.97)	0.01	0.01	0.95 (0.94)	0.03	0.02
	8	0.97 (0.96)	0.02	0.01	0.96 (0.95)	0.02	0.01	0.95 (0.93)	0.03	0.02
	High School	0.99 (0.98)	0.01	0.01	0.97 (0.95)	0.02	0.01	0.93 (0.91)	0.05	0.02
Social Studies	High School	0.98 (0.97)	0.01	0.01	0.96 (0.95)	0.02	0.01	0.92 (0.90)	0.06	0.02

10.4 INTERRATER CONSISTENCY

Chapter 9 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 percent adjacent agreement, the correlation between the first two sets of scores, and the percent of responses that required a third score. This same information is provided at the item level in Appendix E.

**Table 10-4. Summary of Interrater Consistency Statistics
Collapsed across Items by Content Area and Grade**

Content Area	Grade	Number of Items	Number of		Percent		Correlation
			Score Categories	Included Scores	Exact	Adjacent	
English Language Arts	3	12	4	5,756	99.70	0.19	0.99
	4	12	4	7,225	99.39	0.36	0.97
	5	12	4	6,318	99.43	0.28	0.96
	6	12	4	6,920	99.58	0.26	0.98
	7	12	4	7,030	99.66	0.24	0.98
	8	12	4	6,970	99.20	0.53	0.96
	High School	12	4	6,772	99.16	0.59	0.98
Mathematics	3	12	4	5,786	99.52	0.31	0.98
	4	12	4	7,308	99.40	0.33	0.97
	5	12	4	6,400	99.31	0.38	0.96
	6	12	4	7,002	99.27	0.44	0.97
	7	12	4	7,132	99.37	0.42	0.97
	8	12	4	6,902	99.52	0.36	0.98
	High School	12	4	6,734	99.29	0.50	0.98
Science	4	12	4	7,108	99.73	0.20	0.99
	8	12	4	6,818	99.53	0.29	0.97
	High School	12	4	6,830	99.49	0.34	0.98
Social Studies	High School	12	4	6,804	99.27	0.37	0.96

CHAPTER 11 COMPARABILITY (SCALING AND EQUATING)

11.1 COMPARABILITY OF SCORES ACROSS YEARS

In administering the New York State Alternate Assessment (NYSAA), teachers select Alternate Grade Level Indicators (AGLIs), following the Test Blueprints. Use of the 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 Required Components 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, AGLIs, and Scoring Rubrics from year to year. Additionally, comparability is ensured through the scoring process. Teachers use the same Scoring Rubric for a datafolio 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 2008 to establish cut scores for each alternate performance level in English language arts (ELA) and mathematics, Grades 3 through 8 and high school; in science, Grades 4, 8, and high school; and in social studies, Grades 5, 8, and high school. 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 raw score cutpoints for the NYSAAs as established via standard setting are presented in Table 11-1.

Table 11-1. 2012–13 NYSAA: Cut Scores on the Theta Metric and Reporting Scale by Content Area and Grade

Content Area	Grade	Raw Score Cuts			Raw Score	
		Cut 1	Cut 2	Cut 3	Minimum	Maximum
English Language Arts	3	33	40	45	0	48
	4	23	35	46	0	48
	5	33	41	46	0	48
	6	23	39	47	0	48
	7	29	34	44	0	48
	8	30	39	47	0	48
	High School	25	37	47	0	48
Mathematics	3	34	41	46	0	48
	4	28	41	46	0	48
	5	22	34	45	0	48
	6	22	37	43	0	48
	7	32	38	47	0	48
	8	22	35	43	0	48
	High School	32	39	47	0	48
Science	4	29	39	44	0	48
	8	32	37	46	0	48
	High School	27	35	43	0	48
Social Studies	5	20	35	46	0	48
	8	24	36	44	0	48
	High School	33	39	46	0	48

Table F-1 in Appendix F shows performance level distributions for 2013 by content area and grade.

11.1.2 Reported Scores (Cumulative Distributions)

Students' entry scores are calculated based on a combination of their Level of Accuracy and Level of Independence scores for each of the three dates of student performance of the AGLIs in a given entry. The overall score is then the sum of the entry scores. Because of the use of the formula, there may be multiple ways that a student can attain a given total score.

Graphs of the cumulative reported raw score distributions for 2013 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 entitled Steps for Scoring 2012–13 NYSAA Datafolios and Decision Rules for Scoring 2012–13 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 AGLIs describe the content to be included in students' instructional programs for each grade level. The AGLIs are based on the Grade-Level Expectations assessed by the New York State learning standards, but have been reduced in depth and breadth. The AGLIs are designed to follow a developmental continuum of skills that increases across grades. The assessment tasks must align to the AGLIs, which are designed to measure grade-specific content and skills. These assessment tasks and AGLIs, along with Test Blueprints were designed to mirror the developmental continuum reflected in the AGLIs and to ensure that each datafolio builds upon the appropriate knowledge and skills, thereby reflecting the desired continuity across grades.

During administration, the Test Blueprint serves as a guide to the teachers as to how to select AGLIs that are appropriate for a given student. In addition, teachers must select, modify, or create assessment tasks that are aligned with the AGLIs chosen. As with other aspects of the development and administration of the NYSAA, use of the Test Blueprints and the AGLIs ensures that the student is being assessed at a level that is appropriate for his or her individual needs and that the 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 AGLIs. Proficiency across grades, therefore, was expected to follow the developmental continuum established by the AGLIs and, thus, to reflect a higher level of cognition as the grades increased.

CHAPTER 12 VALIDITY

12.1 PROCEDURAL VALIDITY

In order to ensure 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 set 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 contained all the guidelines and specific requirements of the NYSAA; all the forms required to be used in the datafolio; and the Test Blueprints, Alternate Grade Level Indicators (AGLIs), and Sample Assessment Tasks (SATs) for the Required Components for each grade level and content area.
- **Training program video:** The entire Administration Training program that is used with teachers is contained in this video. All AATN Specialists are required to use the video 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 contained slide printouts and Guided Practice activities.

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

- **Steps for Scoring 2012–13 NYSAA Datafolios and Decision Rules for Scoring 2012–13 NYSAA Datafolios:** These are the two main documents used to guide the scoring process for each datafolio (see Appendices B and C).
- **Training program video:** The entire Scoring Training program that is used with Scorers is contained in this video. All SSCs and AATN Specialists are required to use the video 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 four 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, APA, NCME, 1999) 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 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, as well as the guidelines for procedures regarding administration and scoring" (1999, p. 11). In addition to describing the content in detail, content validity evidence must, of course, relate the content to the construct 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" (1999, p. 11).

The New York State (NYS) learning standards provide the framework for the New York State Testing Program, including the NYSAA. These learning 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 NYS learning 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 NYS learning 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 NYS learning standards. The Standard Setting Report documenting the June 2008 standard-setting meeting describes the rigorous procedures that were adhered to in order to ensure that the content-related aspects of the standard-setting maintained a strong substantive alignment with the NYS learning 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 Stakeholders was formed. Their 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, entitled *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 assessment, 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 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 assessment, even with accommodations. The APIs, while based on the 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 the NYSAA, based on several strict criteria, and decided which APIs the students would be assessed on. 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 Act (NCLB).

The following is the calendar of events 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 IDEA of 1997
June 2001	Collect data and report student scores to the public

The Department and its Stakeholders 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 NCLB, states are being held to high levels of student academic achievement, including students with severe cognitive disabilities. The original NYSAA tested students in Grades 4, 8, and in high school in the content areas of English language arts (ELA), mathematics, science/health, and social studies. Based on new testing of grade 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. 2012–13 NYSAA: Revised NYSAA
Grades 3–8 and High School**

<i>Datafolio Component</i>	<i>Grade Equivalents</i>	
	<i>Anchor 4, 8, and High School</i>	<i>Expanded 3, 5, 6, and 7</i>
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 United States Education Department. 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 (i.e., Grade-Level Expectations). However, for these students, Grade-Level Expectations need to be expanded in both breadth and depth. This resulted in the AGLIs contained in the NYSAA Administration Manual: Appendix H—NYSAA Frameworks.

The Department brought together groups of Stakeholders, 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, Stakeholders 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. Chapter 2 of this report contains a more thorough description of the test design and format.

The new NYSAA was first implemented in the late fall of 2006. The administration, which had an 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 United States Education Department. The 2007–08 NYSAA implementation occurred with a full administration period. This administration was based on the refined AGLIs and SATs. The administration again 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 for the 2012–13 administration; therefore, the cut scores established during the June 2008 standard setting remain consistent for each grade level and content area.

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 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 three 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. (1999). *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., and 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

NYSAA Test Blueprint - English Language Arts (ELA) Effective with 2006–07 Administration

REQUIRED COMPONENTS							
Two ELA Key Ideas Must be Assessed at each Grade Level Required Key Ideas Vary by Grade as Marked by an X							
ELA Key Idea ¹	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Reading	X	X	X	X	X	X	X
Writing		X		X		X	X
Listening	X		X		X		
Speaking*	---	---	---	---	---	---	---

*Note: Speaking is not assessed on the general education State assessments.

CHOICE COMPONENTS								
For Each Required Key Idea, There are Two Possible Standards From Which to Draw Allowable Choices of Standard Vary by Grade as Marked by an X Choose One Standard for Each Key Idea from Standards Marked with an X								
Standards	Key Idea	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
1	Reading			X	X	X	X	X
2	Reading	X	X	X	X	X		
3	Reading						X	X
4	Reading	X	X					
1	Writing		X		X		X	X
2	Writing		X		X			
3	Writing						X	X
4	Writing							
1	Listening			X		X		
2	Listening	X		X		X		
3	Listening							
4	Listening	X						

¹Key Ideas are defined on page 2 of the Introduction of the [English Language Arts Core Curriculum \(May 2005\)](#) as the receptive language skills of listening and reading and as the expressive language skills of writing and speaking.

NYSAA Test Blueprint - Mathematics Effective with 2006–07 Administration

REQUIRED COMPONENTS							
Two Mathematics Strands Must be Assessed at each Grade Level Required Strands Vary by Grade as Marked by an X							
MATHEMATICS STRANDS	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Number Sense & Operations	X	X	X	X	X		
Measurement	X	X					
Geometry			X			X	
Algebra				X		X	X
Statistics & Probability					X		X

CHOICE COMPONENTS							
For Each Required Strand, There are Two Possible Bands From Which to Draw Allowable Choices Within Bands Vary by Grade as Marked by an X For Each Required Strand, Choose One of the Bands Marked with an X							
Bands	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
Number Sense & Operations							
Number Systems	X	X	X	X			
Number Theory					X		
Operations	X	X	X	X	X		
Measurement							
Units of Measurement	X	X					
Units/Estimation	X	X					
Geometry							
Geometric Relationships			X			X	
Transformational Geometry						X	
Coordinate Geometry			X				
Algebra							
Variables & Expressions				X		X	X
Equations & Inequalities				X			X
Patterns, Relations & Functions						X	
Statistics & Probability							
Collection of Data							
Organization & Display of Data					X		X
Analysis of Data					X		X

See [Mathematics Core Curriculum \(March 2005\)](#) for further information.

NYSAA Test Blueprint - Science Effective with 2006–07 Administration

REQUIRED COMPONENTS			
Two Standards must be Assessed at each Grade Level as Marked by an X			
Science Standards	Grade 4	Grade 8	High School
1 - Scientific Inquiry	X	X	
4 - Living Environment	X	X	X
4 - Physical Setting/ Earth Science			X

CHOICE COMPONENTS				
For Each Required Standard, There are Two Possible Key Ideas From Which to Draw Key Ideas Vary by Grade as Marked by an X Choose One Key Idea for each Standard from Key Ideas Marked with an X				
Standards	Key Idea	Grade 4	Grade 8	High School*
1 - Scientific Inquiry	1- Develop explanations of natural phenomena	X		
	2- Testing proposed explanations	X	X	
	3- Observations made while testing		X	
4- Living Environment	1- Similarities/differences between living and nonliving things.			X
	3- Changes in organisms over time	X		
	5- Dynamic equilibrium		X	
	7- Human decisions/activities impact			X
4- Physical Setting/ Earth Science	1- Relative motion and perspective			X
	2- Interactions among components of air, water and land	X		X
	3- Particle properties determine observable characteristics of matter and its reactivity		X	

*Note: at the high school level, choices are made within one Standard, i.e., Standard 4. One choice is drawn from the two designated within the Living Environment section of the curriculum and the other choice is drawn from the two designated within the Physical Setting/Earth Science section of the curriculum. See the Core Curricula for Science at <http://www.emsc.nysed.gov/ciai/cores.html#MST>

**NYSAA Test Blueprint - Social Studies (HS only)
Effective with 2010–11 Administration**

REQUIRED COMPONENTS	
Two Standards must be Assessed at the High School Grade Level as Marked by an X	
Social Studies Standards	High School
1 - US History	X
2 - Global History	X

CHOICE COMPONENTS		
For Each Required Standard, There are Two Possible Units From Which to Draw as Marked by an X		
Choose One Unit For Each Standard From Units Marked with an X		
Standards	Units	High School
1- US History	2 - Constitutional Foundations	X
	7 (B) - World in Uncertain Times: 1980-Present	X
2- World History: Global History and Geography	5 - Age of Revolution	X
	8 - Global Connections and Interactions	X

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

APPENDIX B—NYSAA SCORING PROCEDURES

- Follow the steps outlined below to review each NYSAA datafolio.
- If a discrepancy is not addressed in this document, **consult your Table Leader**.
- 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) prior to the Scorer recording the error.

1. Student Demographics, Scorer ID, Scoring Institute Code

a) Is the student demographic information consistent?

Note: Student demographic information must be consistent between the demographic label, Student Page, and Scannable Score Document. If discrepant or if scannable is missing, **consult the Table Leader**. Record Scorer comment **20** at the bottom of the Scorer Worksheet.

b) Do you have a student demographic label?

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) Does the DOB fall within the range indicated on the Student Page for the grade assessed?

- Note: If a DOB is found to be **outside the range specified** for any grade level, **consult your Table Leader**.
- If Measured Progress ProFile™ used - accept the grade level as correct.

If YES →	Mark the grade assessed in the upper right corner of the Scorer Worksheet.	
If NO → <i>Consult the Table Leader</i>	Wrong grade level was assessed.	<u>Record:</u> <ul style="list-style-type: none"> • AGLI code 00099 • “N” for No for all Connections questions for each AGLI within the content areas that should have been assessed • “N” for No Score for each date • Procedural Error comment 1

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

Note: If page two of the Student Page is missing, continue to review and score the assessment.

If YES →	Transcribe any Testing Accommodations to the Scannable Score Document.
If NO →	Continue to review and score the assessment.

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

If YES →	<u>Record:</u> <ul style="list-style-type: none"> • “Y” for Yes for “Was a collegial review of this datafolio conducted?” on the Scannable Score Document.
If NO →	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No • Scorer comment 21, 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.

**Review the documentation to determine the answer to the question for each step.
 The answer will generally be “YES” or “NO”.
 If no instruction for the “YES” answer is provided, proceed to the NEXT part of that step.**

2. Review Sequence of Documentation for each Content Area

ELA, mathematics, science, and social studies. Review entire content area to determine if anything is out of order.
Do not reorganize the datafolio.

a) Are two DSSs present, one for each Required Component?

Identify each DSS by its title (e.g., “Grade 3-ELA” and “Grade 3-ELA Cont’d”).

If NO → <i>Consult the Table Leader</i>	DSS missing.	<u>Record:</u> <ul style="list-style-type: none"> • AGLI code 00099 • “N” for No for each Connections question for the missing AGLI • “N” for No Score for each date • Procedural Error comment 2 Proceed to AGLI that has a DSS or next content area if both DSSs are missing.
	Two DSSs for the same Required Component .	<ul style="list-style-type: none"> • Review and score the first DSS and assessment documentation <u>For the missing AGLI, record:</u> <ul style="list-style-type: none"> • AGLI code 00099 • “N” for No for each Connections question • “N” for No Score for each date • Procedural Error comment 5 Proceed to next content area.

b) Is the documentation for Required Components in order?

Confirm that both Required Components are in the correct order using the titles and the Component box information on the DSSs.

If NO → <i>Consult the Table Leader</i>	Documents are out of order .	<ul style="list-style-type: none"> • Consider documentation that is out of order and score the assessment in the correct order. Do not reorganize datafolio. <u>Record:</u> <ul style="list-style-type: none"> • Scorer comment 22 Proceed to Step 3.
---	-------------------------------------	--

3. Review the DSS: Demographic and Component Information (Start with 1st DSS for 1st AGLI)

a) Is demographic information complete on DSS?

If NO → <i>Consult the Table Leader</i>	Demographic information is discrepant or incomplete .	<ul style="list-style-type: none"> • Transcribe information from the Student Page to the DSS in red ink.
---	--	--

b) Is Choice Component complete on DSS?

If NO → <i>Consult the Table Leader</i>	Choice Component box is not checked on DSS.	<ul style="list-style-type: none"> • Use the AGLI code/text to identify Component in the Frameworks and check appropriate Choice box on DSS in red ink. Proceed to Step 4.
---	--	--

4. AGLI from Grade Level

a) Was Measured Progress ProFile™ used?

If YES →	Accept the AGLI code and text and proceed with Steps 4b and 4c below.
If NO → <i>Consult the Table Leader</i>	Locate the assessed AGLI in the Frameworks using the AGLI code and text recorded on the DSS. Follow questions 4b and 4c below.

b) Is the AGLI indicated on the DSS from the specified Required Component for the student's assessed grade?

If NO → <i>Consult the Table Leader</i>	AGLI code is discrepant or missing from DSS but can be confirmed using AGLI text documented on DSS or VE and Frameworks.	<ul style="list-style-type: none"> Adjust or transcribe AGLI code to DSS in red ink. <u>Record:</u> <ul style="list-style-type: none"> Scorer comment 23a Proceed to next AGLI or content area.
	AGLI for the assessed grade level cannot be confirmed in Frameworks.	<u>Record:</u> <ul style="list-style-type: none"> AGLI code 00099 "N" for No for each Connections question "N" for No Score for each date Procedural Error comment 4 Proceed to next AGLI or content area.
	Recorded AGLI is not from one of the two choice components under this Required Component for the assessed grade.	<u>Record:</u> <ul style="list-style-type: none"> AGLI code 00099 "N" for No for each Connections question "N" for No Score for each date Procedural Error comment 5 Proceed to next AGLI or content area.

c) Does the AGLI text that is documented on the DSS match the text listed in the Frameworks for the confirmed AGLI code?

If YES →	<u>Record:</u> <ul style="list-style-type: none"> AGLI code documented on DSS "Y" for Yes for "AGLI from grade level." Note: Use the "1 st AGLI" space for the first Required Component AGLI and the "2 nd AGLI" space for the second Required Component AGLI.	
If NO → <i>Consult the Table Leader</i>	AGLI text is discrepant or missing but AGLI code is documented , and correct AGLI text can be found in Frameworks.	<ul style="list-style-type: none"> Adjust or transcribe AGLI text on DSS in red ink. <u>Record:</u> <ul style="list-style-type: none"> AGLI code "Y" for Yes for "AGLI from grade level" Scorer comment 23b Proceed to Step 5.
	Discrepant or missing AGLI text cannot be resolved .	<u>Record:</u> <ul style="list-style-type: none"> AGLI code 00099 "N" for No for each Connections question "N" for No Score for each date Procedural Error comment 6 Proceed to next AGLI or content area.

5. Task Connects to AGLI

a) Does the assessment task documented on the DSS clearly connect to the AGLI?

Even if Measured Progress ProFile™ was used or if the teacher documented that the same or a comparable task was used, Scorers must locate the Sample Assessment Task (SAT) in the Frameworks.

If NOT SURE or if TASK IS ORIGINAL – Consult the Table Leader – If ANY of these criteria apply, continue to review and score the assessment:

- Check the Frameworks to see if the task as written appears as an SAT.
- Are the verb/verb phrase and direct object(s) from the AGLI included in the assessment task?
- Does assessment task address the intent of the AGLI and/or the essences of the grade level indicators for the AGLI?
- Does assessment task assess the student on a skill(s) in addition to the intent of the AGLI? (e.g., AGLI states “identify parts of the water cycle” and the assessment task states that “the student will identify parts of the water cycle by labeling each part on a diagram and answer questions about the water cycle” or “the student will identify the parts of a water cycle by creating a water cycle diagram with each of the parts labeled and presenting it to the class”)

If YES →	Assessment task is missing on DSS but can be located on the VE or VE label	<ul style="list-style-type: none"> • If assessment task can be found on VE or VE label, copy to DSS in red ink and confirm that task connects to AGLI. <u>Record</u> <ul style="list-style-type: none"> • Comment 28 or 33
If NO → <i>Consult the Table Leader</i>	Assessment task does not meet criteria - task is not connected to the AGLI.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “Task Connects to AGLI” and remaining Connections question • “N” for No Score for each date of the AGLI • Procedural Error comment 7a Proceed to next AGLI or content area.
	Assessment task is missing and cannot be located on the VE (either evidence itself or VE label).	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “Task Connects to AGLI” and remaining Connections question • “N” for No Score for each date of the AGLI • Procedural Error comment 2 Proceed to next AGLI or content area.

b) Does the assessment task connect to the AGLI by demonstrating 1) any plural contained in the AGLI; 2) any AND, OR, or AND/OR statements contained in the AGLI?

Note: If the assessment task does not contain any plurals or statements, follow “If YES →” directions below.

When the AGLI includes a/an:	Then the ASSESSMENT TASK:
Plural with no parentheses around the “s”.....	MUST demonstrate the plural
Plural with the “s” in parentheses.....	May demonstrate either the singular or plural version of the word
AND statement.....	MUST demonstrate all elements of the AGLI
OR statement.....	May demonstrate one or more of the elements of the AGLI
AND/OR statement.....	May demonstrate one, some or all of the elements of the AGLI

If YES →	<u>Record:</u> <ul style="list-style-type: none"> • “Y” for Yes for “Task connects to AGLI.” Proceed to Step 6. 	
If NO → <i>Consult the Table Leader</i>	AGLI includes a plural without parentheses around the “s” and the plural is not included in the assessment task.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “Task Connects to AGLI” and remaining Connections question • “N” for No Score for each date • Procedural Error comment 7b Proceed to next AGLI or content area.
	AGLI includes an AND statement and not all elements from the AGLI are included in the assessment task.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “Task Connects to AGLI” and remaining Connections question • “N” for No Score for each date • Procedural Error comment 7b Proceed to next AGLI or content area.

6. VE Connects to Task

a) Are two pieces of VE found behind the DSS?

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 MORE THAN TWO PIECES OF VE →	Three or more pieces of VE are included. <i>Note: Do not confuse this with a student work product that is multiple pages or with supporting evidence.</i>	Only the first two pieces of VE can be used to score the assessment. Also, if one or both pieces of evidence for this AGLI are invalid, other evidence cannot be considered in its place. <u>Record:</u> <ul style="list-style-type: none"> • Scorer comment 25 Proceed to Step 6b using only the first two pieces of VE.
If NO → <i>Consult the Table Leader</i>	Only one piece of evidence is found. <i>Note: Review the datafolio to determine if second piece of VE is misplaced. If VE is misplaced, leave it where found, review and score.</i>	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 8 Proceed to next AGLI or content area.

b) Do both pieces of VE connect to the assessment task?

- VE must show how the student demonstrated his or her knowledge, skills, and understanding related to the assessment task. **Note:** If VE is a DCS, it must include step(s), time-segment, or trial information.

If NOT SURE if VE connects to assessment task – Consult the Table Leader – If any of these criteria apply, continue to review and score the assessment:

- Does verifying evidence address the intent of the assessment task? (e.g., the assessment task indicates the student will circle the main idea but the VE shows that the student highlighted the main idea.)
- Does verifying evidence demonstrate the task as written but also includes additional skills? (e.g., the assessment task indicates the student will identify triangles, but the VE shows the student identifying triangles and squares.)

If NO → <i>Consult the Table Leader</i>	If either piece of VE does not connect to the task.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 9a–c Proceed to next AGLI or content area.
	DCS included as VE is missing the steps, time-segment, or trial information.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 15b Proceed to next AGLI or content area.
	Digital video and/or audio malfunctioned or the evidence clip is unable to be located on the DVD and/or CD.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 14b or 14d Proceed to next AGLI or content area.
	DCS includes a single step or one time segment that does not clearly document student performance for the assessed task	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 3a or 3b Proceed to next AGLI or content area.

6. VE Connects to Task (Continued)

- c) Does the VE (in total, both VEs) connect to the assessment task by demonstrating 1) any plural contained in the assessment task; 2) any AND, OR, or AND/OR statements contained in the assessment task; and 3) the vocabulary is consistent with/does not contradict task?

When the Assessment Task includes:	Then the VERIFYING EVIDENCE (in total):
Plural with no parentheses around the “s”....	MUST demonstrate the plural *
Plural with the “s” in parentheses.....	May demonstrate either the singular or plural version of the word to be demonstrated in the VE
AND statement.....	MUST demonstrate <u>all</u> elements of the assessment task *
OR statement.....	May demonstrate one or more of the elements of the assessment task
AND/OR statement.....	May demonstrate one, some or all of the elements of the assessment task
Information on VE (e.g., directions on preprinted worksheet).....	Information on VE is consistent with vocabulary in task
<i>*One piece of VE may contain one element and the other piece of VE may contain the other element.</i>	

If YES →	<u>Record:</u> • “Y” for Yes for “VE connects to task.” Proceed to Step 7.	
If NO → <i>Consult the Table Leader</i>	Assessment task includes a plural without parentheses around the “s.” Upon review of both pieces of VE (in total), they do not satisfy the plural indicated.	<u>Record:</u> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 9d Proceed to next AGLI or content area.
	Assessment task contains an AND statement. Upon review of both pieces of VE (in total), they do not satisfy the AND statement indicated.	<u>Record:</u> • “N” for No for “VE connects to task” • “N” for No Score for each date • Procedural Error comment 9d Proceed to next AGLI or content area.

7. Dates of Student Performance on the DSS

- a) Are three separate dates within the 2012–13 administration period recorded on the DSS (October 1, 2012–February 8, 2013)?

If NO → <i>Consult the Table Leader</i>	One or more dates of student performance within the administration period is missing from DSS, but can be determined from valid VE.	• Transcribe date(s) from VE to the DSS in red ink in chronological order. <u>Record:</u> • Scorer comment 27a
	<i>Note: A valid DCS (see Steps 8a, c, and d) may provide up to three separate dates within the administration period. Any other date(s) on the DCS, not related to assessed date(s), should be ignored. If DCS is valid, use last date(s) recorded on DCS. All other valid VE (see Steps 8a and 8b, e, or f) can provide only one date each.</i>	
	One or more dates of student performance within the administration period cannot be determined from valid VE or one or more dates on DSS are outside the administration period.	<u>Record:</u> • “N” for No Score for date(s) in question • Procedural Error comment 10a or 10b Review the remaining date(s), proceed to next AGLI, or content area.

- b) Are the three dates documented on the DSS in chronological order?

If NO → <i>Consult the Table Leader</i>	Dates are not in chronological order.	• Reorder the dates and student performance data for each date on DSS in red ink. <u>Record:</u> • Scorer comment 27b
--	--	--

7. Dates of Student Performance on the DSS (Continued)

c) Do the dates on each piece of VE correspond to two dates within the administration period on the DSS?

If NO → <i>Consult the Table Leader</i>	Date(s) on the first two pieces of VE are discrepant with the date(s) on DSS , but are within the administration period.	<ul style="list-style-type: none"> Adjust date(s) on the DSS in red ink from the VE. <u>Record:</u> <ul style="list-style-type: none"> Scorer comment 27a or 27c
	First two pieces of VE behind DSS do not confirm two dates of student performance within the administration period.	<u>Record:</u> <ul style="list-style-type: none"> "N" for No Score for date(s) in question Procedural Error comment 10c Accept ratings for the date that does not require evidence Proceed to next AGLI or content area.

8. Valid Evidence: VE and Supporting Evidence (SE)

a) Are the seven required elements clearly documented? (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.

Required Elements			
• Student name	• Content area	• Assessment task	• Level of Independence
• Date of student performance	• AGLI text	• Level of Accuracy	

If YES →	All elements are clearly documented.	Proceed to Steps 8b, c, e, and/or f depending on type of evidence.
If NO → <i>Consult the Table Leader</i>	One or more required elements on the VE and/or VE label is discrepant with the DSS. <i>Note: Use chart below:</i>	<ul style="list-style-type: none"> Adjust the required elements to DSS in red ink. <u>Record:</u> <ul style="list-style-type: none"> Scorer comment 27a or 28a–f Proceed to Steps 8b, c, e, and/or f. <i>Note: Do not make any marks on VE or VE labels</i>
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 → <i>Consult the Table Leader</i>	One or more required elements is missing from VE (VE itself or VE label) or VE label is not affixed to VE.	<u>Record:</u> <ul style="list-style-type: none"> "N" for No Score for that date Procedural Error comment 11a–g Review remaining date(s), proceed to next AGLI, or content area.

b) Is the student work product original? (Student Work Product)

- **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 → <i>Consult the Table Leader</i>	Work product is not original (<i>i.e., photocopies of student responses, correction fluid/tape, black out</i>).	<u>Record:</u> <ul style="list-style-type: none"> "N" for No Score for that date Record Procedural Error comment 12 Review remaining date(s), proceed to next AGLI, or content area.

8. Valid Evidence: VE and SE (Continued)

c) Does the DCS have SE and three dates of data with staff initials for dates that need it? (DCS)

The DCS must have

- a minimum of **three dates**,
- **one piece of SE** for **each date transcribed** to the DSS as VE (acceptable to have up to two dates transcribed as VE, would then have two pieces of SE), and
- have **staff initials** recorded for each date transcribed to the DSS.

If YES →	Continue with Step 8d below; review each submitted piece of SE individually.	
If NO → <i>Consult the Table Leader</i>	Fewer than three dates are documented on the DCS.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for the date(s) transcribed from the DCS to the DSS • Procedural Error comment 15c Review remaining date(s), proceed to next AGLI, or content area.
	SE is missing for a date(s) transcribed to DSS as VE.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for the date(s) transcribed from the DCS to the DSS • Procedural Error comment 15a Review remaining date(s), proceed to next AGLI, or content area.
	Staff initials are missing from DCS for date(s) transcribed to DSS	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date(s) • Procedural Error comment 15d Review remaining date(s), proceed to next AGLI, or content area.

d) Is the SE valid? (Supporting Evidence)

1. Observer Verification Form (OVF)

- Review Step 8a and OVF criteria 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 invalid if:

- any of the seven required elements for valid VE are missing (*Student name, Date of student performance, Content area, AGLI text, Assessment task, Level of Accuracy, Level of Independence*);
- 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 (*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; or
- the observer’s signature date is missing, or **is not the same date task was observed**.

If NO → <i>Consult the Table Leader</i>	OVF is invalid per one or more criteria listed in the bullets above.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 11a–g or 16a–e Review the other piece of SE submitted, remaining date(s) or proceed to next 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> <ul style="list-style-type: none"> • Record Scorer comment 30 Continue to review the other piece of SE submitted or proceed to Step 9.
	Observer’s title is missing from OVF and cannot be confirmed from another OVF in the datafolio.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 16b Review the other piece of SE submitted, remaining date(s) or proceed to next AGLI or content area.

8. Valid Evidence: VE and SE (Continued)

d) Is the SE valid? (Continued)

2. **Student Work Product** - Review Steps 8a and b to determine if student work product is valid SE.
3. **Photographs** - Review Steps 8a and e to determine if photographs are valid SE.
4. **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 → <i>Consult the Table Leader</i>	Student work product, photographs, or digital video or audio clip is invalid per Step criteria .	<u>Record:</u> <ul style="list-style-type: none"> • “N” 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, remaining date(s) or proceed to next AGLI or content area.

e) Are there three photographs with a caption from a single date? (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 → <i>Consult the Table Leader</i>	Fewer than three photographs are submitted of the student performing the task.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 13d Review remaining date(s), proceed to next AGLI, or content area.
	No caption is found.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 13c Review remaining date(s), proceed to next AGLI, or content area.
	No date or multiple dates are found on the evidence.	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for the date • Procedural Error comment 13a or 13b Review remaining date(s), proceed to next AGLI, or content area.

f) Is the digital video and/or audio clip brief and does it contain the seven required elements recorded? (Digital Video/Audio Clip)

Clip must be

- 90 seconds or fewer (excluding markers) and
- contain at least one recorded marker with the seven required elements.

If YES →	Continue to review the other piece of VE submitted or proceed to Step 9.	
If NO → <i>Consult the Table Leader</i>	Clip duration is longer than 90 seconds .	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 14c Review remaining date(s), proceed to next AGLI, or content area.
	All required elements, such as recorded markers, are not recorded on the clip in any manner. <i>Note: VE label on DVD/CD case or box is not acceptable.</i>	<u>Record:</u> <ul style="list-style-type: none"> • “N” for No Score for that date • Procedural Error comment 14a Review remaining date(s), proceed to next AGLI, or content area.

9. Confirm Student Performance Information (Level of Accuracy and Level of Independence Percentages and Rubric Ratings)

- Are the percentages on both pieces of VE calculated correctly and are the rubric ratings (4, 3, 2, or 1) on the DSS recorded correctly?

Review the VE comparing calculations for Level of Accuracy and Level of Independence with the percentages recorded on the DSS. Review rubric ratings corresponding to each percentage. Accept ratings for the date that does not require VE.

If YES →	<p><u>Record:</u></p> <ul style="list-style-type: none"> • Rubric ratings for the Level of Accuracy and Level of Independence <p>Continue to review the other piece of VE submitted or proceed to Step 10.</p>	
If NOT SURE → <i>Consult the Table Leader</i>	<p>Information on the VE contradicts or does not support what is documented for the Level of Accuracy and/or the Level of Independence, and the Scorer cannot clearly see how to correct calculation.</p>	<ul style="list-style-type: none"> • Accept the percentages the teacher documented. <p><u>Record:</u></p> <ul style="list-style-type: none"> • Rubric ratings documented by the teacher • Scorer comment 31
If NO → <i>Consult the Table Leader</i>	<p>Level of Accuracy and/or Level of Independence is missing from the DSS, but is for a date(s) that has valid VE.</p>	<ul style="list-style-type: none"> • Transcribe percentage calculation from the VE to the DSS in red ink and complete the corresponding rubric rating on DSS in red ink. <p><u>Record:</u></p> <ul style="list-style-type: none"> • Rubric ratings • Scorer comment 28e or 28f
	<p><i>Note: A valid DCS (see Steps 8a, c, and d) may provide student performance information for up to three dates within the administration period. If DCS is valid, use the student performance information from the last date(s) recorded on DCS. All other valid VE (see Steps 8a and 8b, e, or f) can only provide student performance information for one date each.</i></p>	
	<p>Level of Accuracy and/or Level of Independence is missing from the DSS for the date that does not require evidence and does not have a valid DCS as VE (see Steps 8a, c, and d).</p>	<p><u>Record:</u></p> <ul style="list-style-type: none"> • "N" for No Score for that date • Procedural Error comment 2 <p>Review remaining date(s), proceed to next AGLI, or content area.</p>
	<p>Level of Accuracy and/or Level of Independence on the VE is discrepant with what is documented on the DSS.</p>	<ul style="list-style-type: none"> • Adjust the percentage calculation and/or rating on the DSS in red ink to match the VE. <p><u>Record:</u></p> <ul style="list-style-type: none"> • Adjusted rubric ratings • Scorer comment 28e or 28f <p><i>Note: Never make changes to VE or VE labels.</i></p>
	<p>Level of Accuracy and/or Level of Independence was incorrectly calculated and the Scorer can clearly see how the percentage calculated can be adjusted.</p> <p><i>Note: if Scorer cannot clearly see how to correct calculation, follow "If NOT SURE →" directions.</i></p>	<ul style="list-style-type: none"> • Adjust the percentage calculation and/or rating on DSS in red ink. <p><u>Record:</u></p> <ul style="list-style-type: none"> • Adjusted rubric ratings • Scorer comment 32

10. Score the 2nd AGLI

Follow Steps 3–9 for the second AGLI from the same content area.

11. Score Mathematics, Science, and Social Studies

Follow Steps 2–10 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).

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

- **AGLI Code, Three Connections Questions, Ratings** - Double check that a **five digit AGLI code** has been recorded; that the **three Connections questions** are bubbled in as “Y” or “N”; and that the **ratings** of 4, 3, 2, 1, N are written in **for each of the three dates for both Accuracy and Independence**.
- **Procedural Error Comments (1–19)** - Double check that a Procedural Error Comment has been recorded on the Scorer Worksheet for each No or No Score.
- **Scorer Comments (20–34) / Positive Feedback Comments (35–41)** - 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.**

13. Complete the Scannable Score Document

Transcribe the following data:	
From the: • Scorer Worksheet	<ul style="list-style-type: none">• AGLI code - 5 digits• Three Connections questions - “Y” for Yes or “N” for No<ul style="list-style-type: none">○ AGLI from grade level○ Task connects to AGLI○ VE connects to task• Ratings (4, 3, 2, 1, N) - Level of Accuracy and Level of Independence
From the: • “Not Tested” form , if applicable	<ul style="list-style-type: none">• Absent• Administrative Error• Not Enrolled• Took Another Assessment• Medically Excused
Confirm you have completed:	
From the: • Student Page	<ul style="list-style-type: none">• 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.
<ul style="list-style-type: none">• Complete the Scannable Score Document for each applicable content area and for any other information as directed by the SSC.• Confirm AGLIs have been recorded correctly -1st AGLI in 1st AGLI space and 2nd AGLI in 2nd AGLI space.	



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

Other Scoring Concerns or Questions Not addressed directly in the Scoring Procedures

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 from 2012-13, or Data Collection Sheet not from NYSAA Administration Manual).
- Task does not connect to the AGLI, **but VE appears to connect to the AGLI.**
- **Photocopies** (either in part or whole) or **correction fluid/tape or black out** is found on assessment documents.
- 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 or supporting evidence** clearly appears to be **homework.**
- VE for ELA is submitted in a **language other than English.**
- Photographic, digital video, or audio evidence appears to **include prerequisite or post-activity steps.**

The following may result in an adjustment within the datafolio.

- **Task description includes a criterion** (e.g., “Student will complete 8 out of 10 problems correctly”) or **prompting** (e.g., “Student will complete task with verbal cue” and Independence is documented as 100%).
- **VE** is a work product that appears to include **prompts toward the correct answer or a format that guides the student directly to the correct answer** (e.g., template).

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

- Use of a “**variety of objects/strategies**” or use of “**concrete objects**” is not clear in the VE.
- **Extra VE or supporting evidence was submitted** beyond the requirements for a specific AGLI.
- **VE label is affixed to incorrect VE** within **same AGLI.**
- Student work product (VE) or VE label is missing assessment task (documented on the DSS), but the **evidence includes directions that restate the assessment task.**
- Chart or calendar is submitted for a **date other than the last date recorded on the chart or calendar.**
- The DCS includes **steps not relevant** to the assessed task, or a **single-step** task is documented on a multi-step DCS.
- **Dates or information printed in the header and/or footer** of documents completed with Measured Progress ProFile™ contradict information recorded on the evidence or VE label.

APPENDIX C—2012–13 SCORING DECISION RULES

Decision Rules for Scoring NYSAA Datafolios

2012–13

(For Table Leaders)

Rule #	Scoring Concern/Question	Decision Rule/Rationale	May come up 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 2011-12, or Data Collection Sheet not from NYSAA Administration Manual).	<p>Incorrect Forms</p> <ul style="list-style-type: none"> If an incorrect Student Page or DSS is used but all assessment requirements can be confirmed, adjust the incorrect information on the form in red ink and score the assessment following the Scoring Procedures. If an incorrect DSS is used and assessment requirements cannot be confirmed, record Alternate Grade Level Indicator (AGLI) code(s) 00099 and “N” for No for each Connections question and “N” for No Score for each date of the AGLI(s). Record Procedural Error comment 17. Continue to next AGLI or content area. <p>Teacher-created Forms</p> <ul style="list-style-type: none"> Teacher created his/her own 2012–13 forms, such as a Data Collection Sheet or VE label. If all requirements are clearly documented, score the assessment following the Scoring Procedures. 	1-8
Assessment Task			
2	Task description includes a criterion (e.g., “Student will complete 8 out of 10 problems correctly”) or prompting (e.g., “Student will complete task with verbal cue” and Independence is documented as 100%).	<p>Criterion in Task Description</p> <ul style="list-style-type: none"> If Level of Accuracy can be determined from verifying evidence (VE), recalculate percentage based on 100% Accuracy and adjust corresponding rating on DSS in red ink for each date. Record the adjusted rating and Scorer comment 33a. Continue to next AGLI or content area. If Level of Accuracy cannot be determined from the VE, adjust Level of Accuracy to 0% and corresponding rating on DSS in red ink for each date. Record the adjusted rating and Scorer comment 33a. Continue to next AGLI or content area. <p>Prompting in Task Description</p> <ul style="list-style-type: none"> If frequency of prompting can be determined from VE, recalculate percentage based on 100% Independence and adjust corresponding rating on DSS in red ink for each date. Record the adjusted rating and Scorer comment 33b. Continue to next AGLI or content area. If frequency of prompting cannot be determined from VE, adjust Level of Independence to 0% and adjust corresponding rating on DSS in red ink for each date. Record the adjusted rating and Scorer comment 33b. Continue to next AGLI or content area. 	5a
3	Task does not connect to the AGLI, but VE appears to connect to the AGLI .	Record “N” for No for “Task connects to AGLI” and remaining Connections question . Record “N” for No Score for each date of the AGLI. Record Procedural Error comment 7 . Continue to next AGLI or content area.	6b-c
4	Photocopies (either in part or whole) or correction fluid/tape or black out is found on assessment documents.	<ul style="list-style-type: none"> Correction fluid/tape or black out found on page numbers, Student Page, or table of contents does not directly impact 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 the datafolio. Score the assessment following the Scoring Procedures. <p>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.</p>	1–8

Verifying Evidence (VE)			
5	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.	<ul style="list-style-type: none"> A student may self-correct on a student work product, which does not require a notation by the teacher. If a teacher-made error is crossed out and corrected but not initialed, score the assessment following the Scoring Procedures. If a teacher-made erasure is confirmed, record “N” for No Score for that date. Record Procedural Error comment 12. Continue to review and score other date(s) for the AGLI following the Scoring Procedures. <p>Note: Documentation made by the teacher does not have to be in permanent ink.</p>	1–9
6	Use of a “variety of objects/strategies” or use of “concrete objects” is not clear in the VE.	It is possible that the use of objects, strategies, or manipulatives will not be clear on a student work product. Unless obvious documentation indicates that the student did not complete the assessment task per the task described, score the assessment following the Scoring Procedures.	6b
7	VE label is affixed to incorrect VE within same AGLI .	Verify required elements and adjust required elements on DSS in red ink , if necessary. Score the assessment following Step 8a and record Scorer comment 28 .	8a
8	Student work product (VE) or VE label is missing assessment task, but the evidence includes directions that restate the assessment task as documented on DSS.	Directions on the student work product restate the assessment task as documented on DSS. Score the assessment following Step 8a. Record Scorer comment 28d .	8a
9	VE or supporting evidence clearly appears to be homework .	<ul style="list-style-type: none"> 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 “N” for No Score for that date. Record Procedural Error comment 18. Continue to score next date. 	7 or 8a–f
10	VE for ELA is submitted in a language other than English .	Record “N” for No Score for that date . Record Procedural Error comment 19 . Continue to score next date.	8b–f
11	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 one date on the DSS. If the date can be verified on the calendar or chart, accept the calendar or chart as evidence for that date. Score the assessment following the Scoring Procedures.	7 or 8b or 8d
12	VE is a work product that appears to include prompts toward the correct answer or a format that guides the student directly to the correct answer (e.g., template).	<p>A template is a direct guide or other format that gives the student the answer is considered a cue or prompt and impacts the student’s Level of Independence.</p> <ul style="list-style-type: none"> If VE appears to include prompts of correct answers or a direct guide, recalculate Level of Independence percentage based on items that were completed without the prompts or guide divided by the total number of items; if a prompt or guide is applied to a work product as a whole, then adjust Level of Independence to 0%. Adjust the percentage and corresponding rating on the DSS in red ink for that date for Level of Independence. Record the rating for Level of Accuracy and adjusted rating for Level of Independence. Record Scorer comment 34. Continue to score the next date. <p>Examples:</p> <ul style="list-style-type: none"> AGLI and task are assessing putting events in sequence: VE is a sequencing worksheet that contains three boxes labeled “First,” “Next,” “Last”; the student response choices are pictures that contain the words “First,” “Next,” “Last.” AGLI and task are assessing locating New York: VE is a map of the northeast with all of the state names typed, except for New York which is handwritten; the directions state “Locate New York, and mark it.” AGLI and task are assessing recognizing parts of a plant: VE has a picture of the plant that has the structures next to each one with an arrow pointing to the structure, and the response choices on the worksheet are the exact same pictures of the structures. AGLI and task are assessing identifying congruent shapes: VE has directions at the top that say “Congruent shapes are the same size and shape. Circle the shapes below that are congruent.” 	8b or 8d or 9

13	Photographic, digital video, or audio evidence appears to include prerequisite or post-activity steps .	<ul style="list-style-type: none"> • All of the requirements for VE are met and the additional requirements for photographic, digital video, or audio evidence are met. Accept what is documented by the teacher and score the assessment following the Scoring Procedures. • If requirements for VE and the other requirements for photographic, digital video, or audio evidence are not met, record "N" for No Score for that date. Record Procedural Error comment 13 or 14. Continue to score the next date. 	8b, 8e, or 8f or 9
14	The DCS includes steps not relevant to the assessed task, a single-step task, or single time segment is documented on a DCS.	<ul style="list-style-type: none"> • 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 (all steps listed on the DCS are scored, unless the teacher clearly indicates otherwise). • If a single-step task is documented on a multi-step DCS, score the assessment following the Scoring Procedures as documented. • If a single time segment is documented on a DCS, score the assessment following the Scoring procedures as documented. 	8c or 9
Dates			
15	Dates or information printed in the 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.	1–9

APPENDIX D—SUBGROUP RELIABILITY

**Table D-1. 2012–13 NYSAA: Subgroup Reliabilities—
English Language Arts**

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
3	All Students	2,748	48	45.24	5.84	0.86	2.22
	Male	1,893	48	45.24	5.86	0.85	2.25
	Female	855	48	45.26	5.78	0.86	2.14
	American Indian/Alaskan Native	21	48	45.24	5.81	0.76	2.83
	Black	688	48	45.69	5.43	0.87	1.98
	Asian	140	48	45.76	4.90	0.89	1.66
	Hispanic	775	48	45.25	6.40	0.89	2.16
	White	1,086	48	44.85	5.81	0.82	2.44
	Native Hawaiian/Other Pacific Islander	11	48	47.55	0.82	0.41	0.63
Multi	27	48	45.85	4.57	0.89	1.52	
4	All Students	3,019	48	45.26	5.81	0.84	2.34
	Male	2,066	48	45.29	5.72	0.83	2.38
	Female	953	48	45.21	5.99	0.86	2.28
	American Indian/Alaskan Native	34	48	47.35	2.06	0.82	0.87
	Black	750	48	45.69	5.49	0.83	2.26
	Asian	165	48	45.13	5.39	0.78	2.50
	Hispanic	794	48	45.55	6.09	0.83	2.52
	White	1,238	48	44.78	5.90	0.84	2.36
	Native Hawaiian/Other Pacific Islander	9	48				
Multi	29	48	46.28	2.72	0.57	1.78	
5	All Students	3,211	48	45.55	5.32	0.83	2.17
	Male	2,231	48	45.58	5.34	0.84	2.15
	Female	980	48	45.47	5.29	0.83	2.21
	American Indian/Alaskan Native	28	48	45.07	6.38	0.77	3.07
	Black	807	48	45.98	4.94	0.84	1.97
	Asian	147	48	45.68	4.80	0.82	2.04
	Hispanic	807	48	45.92	5.06	0.81	2.22
	White	1,377	48	45.06	5.69	0.84	2.31
	Native Hawaiian/Other Pacific Islander	12	48	46.58	3.20	0.69	1.78
Multi	33	48	45.39	5.02	0.87	1.81	
6	All Students	3,295	48	45.25	5.92	0.85	2.30
	Male	2,234	48	45.39	5.70	0.84	2.28
	Female	1,061	48	44.97	6.36	0.87	2.32
	American Indian/Alaskan Native	22	48	44.64	7.10	0.68	4.03
	Black	858	48	45.54	5.87	0.87	2.08
	Asian	167	48	44.51	6.86	0.85	2.62
	Hispanic	876	48	45.80	5.35	0.86	2.02
	White	1,339	48	44.79	6.15	0.83	2.54
	Native Hawaiian/Other Pacific Islander	5	48				
Multi	28	48	46.89	2.33	0.70	1.29	
7	All Students	3,155	48	45.23	6.01	0.83	2.48
	Male	2,154	48	45.21	6.05	0.83	2.53
	Female	1,001	48	45.29	5.92	0.84	2.38
	American Indian/Alaskan Native	22	48	46.41	5.30	0.65	3.11
	Black	852	48	45.30	6.45	0.82	2.72

continued

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
7	Asian	167	48	45.35	6.10	0.87	2.19
	Hispanic	784	48	45.35	6.13	0.83	2.50
	White	1,308	48	45.09	5.62	0.82	2.36
	Native Hawaiian/Other Pacific Islander	6	48				
	Multi	16	48	45.00	6.45	0.63	3.93
8	All Students	3,111	48	45.06	5.81	0.85	2.26
	Male	2,129	48	44.96	5.88	0.84	2.34
	Female	982	48	45.26	5.67	0.86	2.09
	American Indian/Alaskan Native	16	48	47.31	1.78	0.77	0.85
	Black	783	48	45.02	6.05	0.87	2.16
	Asian	152	48	44.58	6.25	0.87	2.21
	Hispanic	812	48	45.56	5.54	0.88	1.95
	White	1,325	48	44.77	5.82	0.82	2.50
	Native Hawaiian/Other Pacific Islander	6	48				
	Multi	17	48	46.88	2.57	0.86	0.96
High School	All Students	2,968	48	44.22	6.43	0.84	2.56
	Male	1,918	48	44.25	6.39	0.85	2.51
	Female	1,050	48	44.16	6.52	0.84	2.65
	American Indian/Alaskan Native	21	48	45.14	4.46	0.84	1.76
	Black	752	48	44.75	6.03	0.85	2.32
	Asian	173	48	44.01	6.78	0.85	2.65
	Hispanic	683	48	44.23	6.59	0.84	2.63
	White	1,322	48	43.94	6.52	0.84	2.63
	Native Hawaiian/Other Pacific Islander	2	48				
	Multi	15	48	42.67	8.13	0.71	4.34

**Table D-2. 2012–13 NYSAA: Subgroup Reliabilities—
Mathematics**

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
3	All Students	2,749	48	45.33	5.28	0.86	2.00
	Male	1,894	48	45.50	5.12	0.85	2.01
	Female	855	48	44.97	5.62	0.87	2.02
	American Indian/Alaskan Native	21	48	44.67	6.04	0.76	2.95
	Black	689	48	45.84	5.03	0.87	1.82
	Asian	140	48	45.54	5.19	0.89	1.72
	Hispanic	773	48	45.74	5.21	0.88	1.83
	White	1,088	48	44.66	5.49	0.83	2.25
	Native Hawaiian/Other Pacific Islander	11	48	47.00	2.41	0.88	0.83
	Multi	27	48	46.41	2.85	0.69	1.58
4	All Students	3,015	48	45.33	5.38	0.85	2.05
	Male	2,063	48	45.32	5.42	0.85	2.10
	Female	952	48	45.34	5.30	0.86	1.95
	American Indian/Alaskan Native	34	48	44.44	7.16	0.84	2.85
	Black	748	48	45.78	5.14	0.83	2.13

continued

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
4	Asian	166	48	45.04	5.78	0.84	2.28
	Hispanic	795	48	46.10	4.66	0.85	1.78
	White	1,234	48	44.58	5.79	0.86	2.18
	Native Hawaiian/Other Pacific Islander	9	48				
	Multi	29	48	46.62	2.37	0.69	1.31
5	All Students	3,202	48	45.71	5.06	0.85	1.94
	Male	2,224	48	45.74	5.11	0.86	1.94
	Female	978	48	45.63	4.93	0.85	1.92
	American Indian/Alaskan Native	28	48	45.75	5.77	0.72	3.07
	Black	808	48	46.05	4.70	0.85	1.83
	Asian	147	48	45.82	4.28	0.83	1.76
	Hispanic	807	48	46.19	4.54	0.84	1.84
	White	1,367	48	45.20	5.57	0.86	2.08
	Native Hawaiian/Other Pacific Islander	12	48	46.67	3.20	0.72	1.68
	Multi	33	48	45.73	4.61	0.85	1.79
6	All Students	3,298	48	44.72	6.72	0.88	2.28
	Male	2,234	48	44.77	6.64	0.88	2.29
	Female	1,064	48	44.61	6.88	0.89	2.27
	American Indian/Alaskan Native	22	48	46.59	2.15	0.58	1.40
	Black	861	48	44.84	7.02	0.90	2.17
	Asian	167	48	44.91	6.49	0.91	1.92
	Hispanic	877	48	44.81	7.16	0.91	2.17
	White	1,339	48	44.48	6.34	0.85	2.44
	Native Hawaiian/Other Pacific Islander	5	48				
	Multi	27	48	46.30	3.38	0.81	1.47
7	All Students	3,158	48	45.23	5.79	0.84	2.28
	Male	2,155	48	45.22	5.84	0.83	2.40
	Female	1,003	48	45.26	5.70	0.87	2.07
	American Indian/Alaskan Native	22	48	45.73	3.55	0.67	2.05
	Black	848	48	45.69	5.46	0.82	2.30
	Asian	168	48	45.02	6.39	0.88	2.20
	Hispanic	787	48	45.40	5.88	0.86	2.22
	White	1,311	48	44.85	5.92	0.84	2.36
	Native Hawaiian/Other Pacific Islander	6	48				
	Multi	16	48	45.63	3.20	0.62	1.99
8	All Students	3,111	48	44.81	6.46	0.85	2.51
	Male	2,128	48	44.83	6.46	0.85	2.54
	Female	983	48	44.78	6.47	0.86	2.45
	American Indian/Alaskan Native	16	48	44.94	5.57	0.84	2.19
	Black	784	48	44.77	6.79	0.84	2.70
	Asian	153	48	45.06	5.74	0.80	2.55
	Hispanic	812	48	45.11	6.51	0.89	2.19
	White	1,323	48	44.62	6.32	0.83	2.58
	Native Hawaiian/Other Pacific Islander	6	48				
	Multi	17	48	45.35	6.04	0.73	3.15
High School	All Students	2,949	48	44.47	6.49	0.85	2.53
	Male	1,905	48	44.37	6.66	0.85	2.59
	Female	1,044	48	44.64	6.16	0.85	2.42

continued

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
High School	American Indian/Alaskan Native	21	48	44.52	6.84	0.79	3.15
	Black	748	48	44.96	6.12	0.85	2.40
	Asian	172	48	44.10	6.71	0.85	2.60
	Hispanic	682	48	44.63	6.47	0.86	2.43
	White	1,309	48	44.16	6.64	0.84	2.62
	Native Hawaiian/Other Pacific Islander	2	48				
	Multi	15	48	43.20	8.68	0.78	4.11

**Table D-3. 2012–13 NYSAA: Subgroup Reliabilities—
Science**

Grade	Group	Number of Students	Raw Score			Alpha	SEM
			Maximum	Mean	Standard Deviation		
4	All Students	3,012	48	45.59	5.66	0.84	2.30
	Male	2,059	48	45.53	5.67	0.83	2.35
	Female	953	48	45.74	5.62	0.85	2.18
	American Indian/Alaskan Native	34	48	46.62	4.48	0.74	2.27
	Black	748	48	45.72	5.70	0.83	2.37
	Asian	166	48	45.82	4.58	0.73	2.36
	Hispanic	795	48	45.82	5.89	0.79	2.72
	White	1,231	48	45.26	5.70	0.85	2.20
	Native Hawaiian/Other Pacific Islander	9	48				
	Multi	29	48	47.21	1.45	0.56	0.97
8	All Students	3,100	48	44.61	6.77	0.84	2.69
	Male	2,122	48	44.61	6.82	0.85	2.67
	Female	978	48	44.60	6.68	0.83	2.74
	American Indian/Alaskan Native	16	48	46.75	2.32	0.55	1.56
	Black	782	48	44.39	7.26	0.84	2.89
	Asian	152	48	44.02	7.81	0.86	2.96
	Hispanic	806	48	44.71	7.08	0.87	2.52
	White	1,321	48	44.70	6.21	0.82	2.61
	Native Hawaiian/Other Pacific Islander	6	48				
	Multi	17	48	46.06	2.44	0.37	1.93
High School	All Students	2,964	48	44.85	6.02	0.84	2.40
	Male	1,915	48	44.92	5.93	0.83	2.41
	Female	1,049	48	44.74	6.18	0.85	2.39
	American Indian/Alaskan Native	21	48	45.67	3.79	0.74	1.93
	Black	752	48	45.56	5.62	0.84	2.28
	Asian	172	48	45.24	4.80	0.84	1.93
	Hispanic	683	48	44.39	6.70	0.86	2.46
	White	1,319	48	44.65	6.02	0.83	2.48
	Native Hawaiian/Other Pacific Islander	2	48				
	Multi	15	48	43.07	5.76	0.84	2.32

**Table D-4. 2012–13 NYSAA: Subgroup Reliabilities—
Social Studies**

<i>Grade</i>	<i>Group</i>	<i>Number of Students</i>	<i>Raw Score</i>			<i>Alpha</i>	<i>SEM</i>
			<i>Maximum</i>	<i>Mean</i>	<i>Standard Deviation</i>		
	All Students	2,970	48	45.08	5.85	0.84	2.33
	Male	1,920	48	45.00	5.93	0.83	2.42
	Female	1,050	48	45.23	5.69	0.85	2.17
High School	American Indian/Alaskan Native	21	48	43.48	7.63	0.88	2.69
	Black	752	48	45.60	5.44	0.85	2.11
	Asian	172	48	45.02	5.68	0.85	2.23
	Hispanic	685	48	44.90	6.23	0.87	2.21
	White	1,323	48	44.92	5.85	0.82	2.50
	Native Hawaiian/Other Pacific Islander	2	48				
	Multi	15	48	44.40	5.58	0.86	2.12

APPENDIX E—INTERRATER CONSISTENCY

**Table E-1. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
English Language Arts Grade 3**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	483	99.38	0.41	0.96
ACC1_2	4	483	99.79	0.00	0.98
ACC1_3	4	482	99.38	0.41	0.96
ACC2_1	4	477	99.79	0.00	0.98
ACC2_2	4	477	100.00	0.00	1.00
ACC2_3	4	476	100.00	0.00	1.00
IND1_1	4	483	99.38	0.62	0.99
IND1_2	4	483	99.79	0.00	0.97
IND1_3	4	482	99.17	0.62	0.97
IND2_1	4	477	99.79	0.21	1.00
IND2_2	4	477	100.00	0.00	1.00
IND2_3	4	476	100.00	0.00	1.00

**Table E-2. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
English Language Arts Grade 4**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	612	99.67	0.00	0.97
ACC1_2	4	611	99.02	0.82	0.96
ACC1_3	4	613	99.67	0.16	0.96
ACC2_1	4	593	98.99	0.84	0.98
ACC2_2	4	591	99.32	0.68	0.99
ACC2_3	4	593	98.82	0.84	0.93
IND1_1	4	612	99.35	0.16	0.94
IND1_2	4	611	99.35	0.16	0.94
IND1_3	4	613	99.84	0.00	0.98
IND2_1	4	593	99.49	0.17	0.99
IND2_2	4	590	99.66	0.17	0.99
IND2_3	4	593	99.49	0.34	0.98

**Table E-3. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
English Language Arts Grade 5**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	535	99.44	0.19	0.96
ACC1_2	4	534	99.63	0.19	0.93
ACC1_3	4	534	99.44	0.19	0.87
ACC2_1	4	518	99.81	0.19	1.00
ACC2_2	4	519	99.81	0.19	0.99
ACC2_3	4	519	99.42	0.39	0.97

continued

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
IND1_1	4	535	98.50	0.93	0.96
IND1_2	4	534	99.63	0.19	0.97
IND1_3	4	534	99.25	0.56	0.97
IND2_1	4	518	99.42	0.39	0.98
IND2_2	4	519	99.42	0.00	0.93
IND2_3	4	519	99.42	0.00	0.94

Table E-4. 2012–13 NYSA: Item-Level Interrater Consistency Statistics—English Language Arts Grade 6

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	586	100.00	0.00	1.00
ACC1_2	4	586	99.49	0.34	0.98
ACC1_3	4	585	99.83	0.00	0.97
ACC2_1	4	568	99.47	0.18	0.97
ACC2_2	4	566	99.29	0.71	0.98
ACC2_3	4	569	99.47	0.35	0.97
IND1_1	4	586	99.66	0.00	0.97
IND1_2	4	586	99.66	0.17	0.98
IND1_3	4	585	99.83	0.00	0.99
IND2_1	4	568	99.30	0.53	0.99
IND2_2	4	566	99.47	0.53	0.99
IND2_3	4	569	99.47	0.35	0.97

Table E-5. 2012–13 NYSA: Item-Level Interrater Consistency Statistics—English Language Arts Grade 7

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	596	99.50	0.34	0.98
ACC1_2	4	597	99.83	0.17	1.00
ACC1_3	4	591	99.83	0.17	1.00
ACC2_1	4	577	99.13	0.69	0.98
ACC2_2	4	578	99.65	0.17	0.98
ACC2_3	4	576	99.48	0.52	0.99
IND1_1	4	596	99.66	0.17	0.97
IND1_2	4	597	99.83	0.17	1.00
IND1_3	4	591	99.66	0.17	0.96
IND2_1	4	577	99.83	0.17	1.00
IND2_2	4	578	99.83	0.00	0.96
IND2_3	4	576	99.65	0.17	0.95

**Table E-6. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
English Language Arts Grade 8**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	582	99.83	0.17	1.00
ACC1_2	4	581	99.48	0.52	0.99
ACC1_3	4	581	99.31	0.69	0.98
ACC2_1	4	580	98.97	0.69	0.97
ACC2_2	4	581	99.31	0.69	0.99
ACC2_3	4	580	99.66	0.34	0.99
IND1_1	4	582	99.14	0.34	0.96
IND1_2	4	581	98.45	0.69	0.90
IND1_3	4	581	98.11	1.20	0.88
IND2_1	4	580	99.48	0.52	0.99
IND2_2	4	581	99.31	0.34	0.95
IND2_3	4	580	99.31	0.17	0.93

**Table E-7. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
English Language Arts High School**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	570	98.77	0.70	0.97
ACC1_2	4	571	99.47	0.35	0.98
ACC1_3	4	572	99.30	0.35	0.95
ACC2_1	4	558	99.28	0.72	0.99
ACC2_2	4	558	98.75	1.08	0.97
ACC2_3	4	557	99.10	0.54	0.96
IND1_1	4	570	98.42	1.40	0.98
IND1_2	4	571	99.12	0.53	0.97
IND1_3	4	572	99.13	0.70	0.98
IND2_1	4	558	99.46	0.36	0.99
IND2_2	4	558	99.64	0.00	0.99
IND2_3	4	557	99.46	0.36	0.98

**Table E-8. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 3**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	482	99.59	0.41	0.99
ACC1_2	4	483	99.59	0.41	0.99
ACC1_3	4	481	99.38	0.62	0.98
ACC2_1	4	482	99.79	0.00	0.98
ACC2_2	4	483	99.59	0.21	0.98
ACC2_3	4	482	99.79	0.00	0.98
IND1_1	4	482	98.96	0.62	0.97

continued

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
IND1_2	4	483	99.59	0.21	0.97
IND1_3	4	481	99.79	0.00	0.98
IND2_1	4	482	98.96	0.62	0.96
IND2_2	4	483	99.38	0.41	0.97
IND2_3	4	482	99.79	0.21	1.00

**Table E-9. 2012–13 NYSA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 4**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	616	99.84	0.00	0.99
ACC1_2	4	613	99.84	0.16	1.00
ACC1_3	4	615	99.67	0.33	0.99
ACC2_1	4	605	99.34	0.33	0.98
ACC2_2	4	603	99.50	0.17	0.95
ACC2_3	4	602	99.17	0.50	0.93
IND1_1	4	616	99.35	0.32	0.97
IND1_2	4	613	99.35	0.65	0.99
IND1_3	4	615	99.67	0.00	0.97
IND2_1	4	605	98.68	0.83	0.96
IND2_2	4	603	99.17	0.50	0.97
IND2_3	4	602	99.17	0.17	0.94

**Table E-10. 2012–13 NYSA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 5**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	539	99.63	0.19	0.99
ACC1_2	4	539	99.44	0.56	0.99
ACC1_3	4	540	99.81	0.19	0.99
ACC2_1	4	528	98.86	0.76	0.94
ACC2_2	4	526	99.05	0.76	0.94
ACC2_3	4	528	99.24	0.57	0.92
IND1_1	4	539	99.26	0.37	0.99
IND1_2	4	539	99.26	0.00	0.94
IND1_3	4	540	99.44	0.00	0.95
IND2_1	4	527	99.62	0.00	0.98
IND2_2	4	527	98.86	0.57	0.93
IND2_3	4	528	99.24	0.57	0.97

**Table E-11. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 6**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	582	98.97	0.69	0.98
ACC1_2	4	582	99.31	0.34	0.97
ACC1_3	4	582	99.66	0.34	0.99
ACC2_1	4	585	99.32	0.34	0.98
ACC2_2	4	585	99.15	0.68	0.98
ACC2_3	4	585	99.32	0.68	0.99
IND1_1	4	582	99.14	0.52	0.96
IND1_2	4	582	99.14	0.34	0.93
IND1_3	4	582	99.31	0.17	0.94
IND2_1	4	585	99.15	0.68	0.98
IND2_2	4	585	99.49	0.17	0.98
IND2_3	4	585	99.32	0.34	0.96

**Table E-12. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 7**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	600	99.67	0.33	1.00
ACC1_2	4	599	99.50	0.33	0.98
ACC1_3	4	595	99.50	0.17	0.93
ACC2_1	4	591	99.49	0.51	0.99
ACC2_2	4	591	99.49	0.34	0.97
ACC2_3	4	590	99.49	0.34	0.95
IND1_1	4	600	99.33	0.33	0.97
IND1_2	4	599	99.17	0.33	0.97
IND1_3	4	595	99.66	0.00	0.97
IND2_1	4	591	98.65	1.02	0.95
IND2_2	4	591	99.15	0.68	0.96
IND2_3	4	590	99.32	0.68	0.98

**Table E-13. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Mathematics Grade 8**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	584	99.32	0.68	0.99
ACC1_2	4	584	99.32	0.68	0.98
ACC1_3	4	582	99.66	0.17	0.95
ACC2_1	4	568	99.12	0.88	0.99
ACC2_2	4	566	99.29	0.71	0.99
ACC2_3	4	567	99.65	0.18	0.97
IND1_1	4	584	99.66	0.34	1.00

continued

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
IND1_2	4	584	99.83	0.00	0.98
IND1_3	4	582	99.48	0.17	0.94
IND2_1	4	568	99.30	0.53	0.98
IND2_2	4	566	99.65	0.00	0.96
IND2_3	4	567	100.00	0.00	1.00

**Table E-14. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Mathematics High School**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	562	98.93	0.71	0.96
ACC1_2	4	563	98.40	1.42	0.97
ACC1_3	4	562	99.11	0.53	0.96
ACC2_1	4	561	99.64	0.18	0.99
ACC2_2	4	559	99.28	0.54	0.96
ACC2_3	4	560	100.00	0.00	1.00
IND1_1	4	562	98.58	0.89	0.97
IND1_2	4	563	98.58	0.89	0.96
IND1_3	4	562	99.47	0.36	0.98
IND2_1	4	561	99.64	0.36	1.00
IND2_2	4	559	99.82	0.18	1.00
IND2_3	4	560	100.00	0.00	1.00

**Table E-15. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Science Grade 4**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	586	100.00	0.00	1.00
ACC1_2	4	587	99.66	0.17	0.97
ACC1_3	4	587	99.49	0.51	0.98
ACC2_1	4	599	99.67	0.33	0.99
ACC2_2	4	598	99.83	0.00	0.96
ACC2_3	4	597	99.66	0.17	0.96
IND1_1	4	586	100.00	0.00	1.00
IND1_2	4	587	99.66	0.34	0.99
IND1_3	4	587	100.00	0.00	1.00
IND2_1	4	599	99.67	0.17	0.99
IND2_2	4	598	99.50	0.33	0.97
IND2_3	4	597	99.66	0.34	0.99

**Table E-16. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Science Grade 8**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	577	99.31	0.52	0.99
ACC1_2	4	575	99.30	0.35	0.96
ACC1_3	4	574	99.48	0.52	0.99
ACC2_1	4	561	99.47	0.53	0.99
ACC2_2	4	562	99.82	0.18	1.00
ACC2_3	4	560	99.82	0.00	0.96
IND1_1	4	577	98.96	0.69	0.96
IND1_2	4	575	99.13	0.35	0.92
IND1_3	4	574	99.65	0.00	0.96
IND2_1	4	561	99.82	0.18	1.00
IND2_2	4	562	99.82	0.00	0.97
IND2_3	4	560	99.82	0.18	1.00

**Table E-17. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Science High School**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	574	99.65	0.35	1.00
ACC1_2	4	574	99.83	0.17	1.00
ACC1_3	4	575	99.65	0.17	0.96
ACC2_1	4	564	98.94	1.06	0.99
ACC2_2	4	564	99.65	0.18	0.98
ACC2_3	4	564	99.11	0.53	0.91
IND1_1	4	574	100.00	0.00	1.00
IND1_2	4	574	99.48	0.35	0.99
IND1_3	4	575	99.13	0.17	0.94
IND2_1	4	564	98.94	0.71	0.96
IND2_2	4	564	99.47	0.35	0.97
IND2_3	4	564	100.00	0.00	1.00

**Table E-18. 2012–13 NYSAA: Item-Level Interrater Consistency Statistics—
Social Studies High School**

<i>Item Number</i>	<i>Number of</i>		<i>Percent</i>		<i>Correlation</i>
	<i>Score Categories</i>	<i>Responses Scored Twice</i>	<i>Exact</i>	<i>Adjacent</i>	
ACC1_1	4	573	99.30	0.35	0.97
ACC1_2	4	573	99.48	0.35	0.97
ACC1_3	4	572	99.48	0.35	0.96
ACC2_1	4	562	98.93	0.53	0.93
ACC2_2	4	561	99.11	0.53	0.93
ACC2_3	4	561	98.75	0.36	0.86
IND1_1	4	573	99.13	0.70	0.99
IND1_2	4	573	99.48	0.17	0.97
IND1_3	4	572	100.00	0.00	1.00
IND2_1	4	562	99.29	0.36	0.97
IND2_2	4	561	99.11	0.36	0.94
IND2_3	4	561	99.11	0.36	0.93

APPENDIX F—PERFORMANCE LEVEL DISTRIBUTIONS

**Table F-1. 2012–13 NYSAA: Performance Level Distributions
by Content Area and Grade**

Content Area	Grade	Performance Level	Percent at Level		
			2012–13	2011–12	2010–11
English Language Arts	3	4	79.22	76.04	72.47
		3	10.08	11.65	13.26
		2	5.06	5.49	6.42
		1	5.64	6.83	7.85
	4	4	75.19	70.77	70.29
		3	12.62	14.51	14.68
		2	6.43	8.22	8.71
		1	5.76	6.49	6.32
	5	4	69.95	64.72	63.43
		3	22.39	24.50	26.35
		2	4.30	5.66	6.30
		1	3.36	5.11	3.92
	6	4	74.72	69.62	66.77
		3	13.20	15.93	17.66
		2	8.41	9.87	11.05
		1	3.67	4.58	4.52
	7	4	84.72	81.88	80.23
		3	7.67	9.82	10.65
		2	6.50	7.42	7.97
		1	1.11	0.88	1.14
8	4	82.64	81.83	77.88	
	3	10.29	12.04	14.05	
	2	6.24	5.34	7.06	
	1	0.84	0.79	1.01	
High School	4	76.99	80.01	75.74	
	3	14.22	12.34	14.81	
	2	4.38	4.31	5.46	
	1	4.41	3.34	3.99	
Mathematics	3	4	73.19	68.73	67.18
		3	20.81	22.19	23.62
		2	5.38	7.81	7.89
		1	0.62	1.27	1.31
	4	4	67.89	64.83	62.33
		3	22.45	24.09	24.86
		2	8.79	10.41	11.74
		1	0.86	0.67	1.08
	5	4	71.58	69.48	66.59
		3	22.02	22.03	24.03
		2	4.81	6.38	6.43
		1	1.59	2.11	2.95
	6	4	76.53	73.95	70.77
		3	15.19	17.12	19.37
		2	6.49	7.49	8.24
		1	1.79	1.44	1.63
	7	4	68.14	64.35	60.99
		3	23.27	24.64	28.22

continued

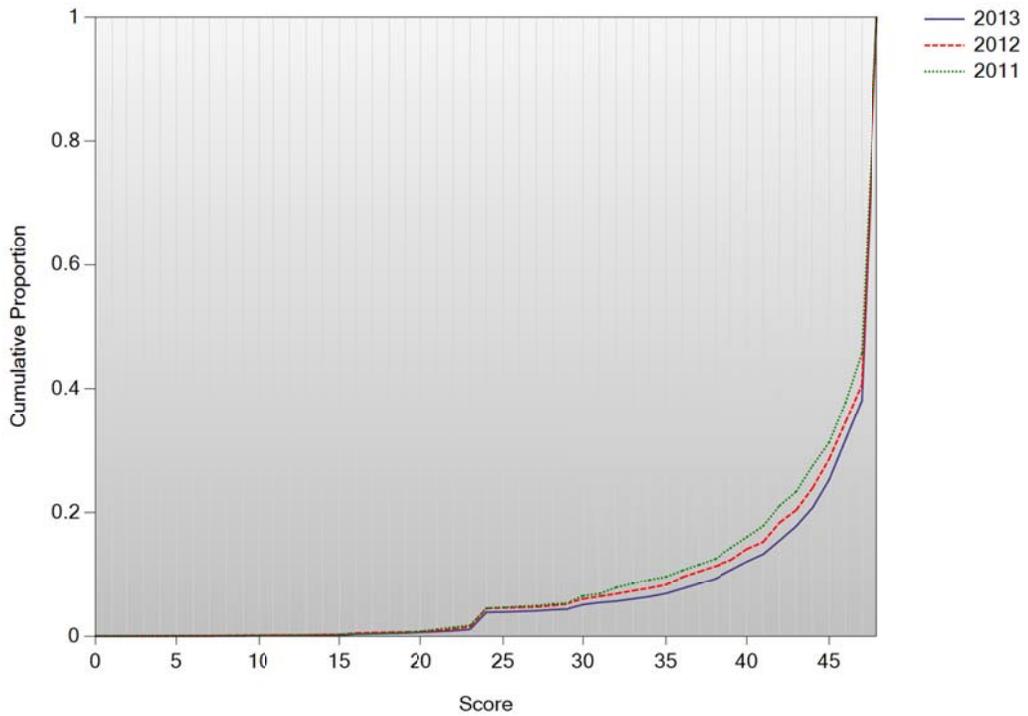
Content Area	Grade	Performance Level	Percent at Level			
			2012–13	2011–12	2010–11	
Mathematics	7	2	3.42	4.91	4.72	
		1	5.16	6.09	6.07	
	8	4	66.60	65.75	61.45	
		3	22.24	24.56	27.16	
		2	4.79	3.73	4.88	
		1	6.36	5.95	6.51	
	High School	4	68.26	69.65	63.99	
		3	23.02	23.15	26.21	
		2	7.90	6.48	9.04	
		1	0.81	0.71	0.77	
	Science	4	4	84.86	82.41	80.29
			3	9.43	11.53	12.07
2			1.86	2.27	2.75	
1			3.85	3.78	4.90	
8		4	79.52	79.62	75.89	
		3	8.42	9.66	10.52	
		2	6.03	5.82	7.38	
		1	6.03	4.90	6.21	
High School		4	78.95	80.49	76.99	
		3	13.12	12.88	14.55	
		2	6.48	5.54	7.16	
		1	1.45	1.09	1.30	
Social Studies	High School	4	72.32	72.35	67.87	
		3	17.61	18.26	19.77	
		2	3.94	4.23	5.83	
		1	6.13	5.16	6.53	

4 = Meeting standards with distinction; 3 = Meeting standards; 2 = Partially meeting standards; 1 = Not meeting standards.

APPENDIX G—CUMULATIVE DISTRIBUTION GRAPHS

Figure G-1. 2012–13 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 3 Bottom: English Language Arts Grade 4

Cumulative Score Distributions: English Language Arts Grade 3



Cumulative Score Distributions: English Language Arts Grade 4

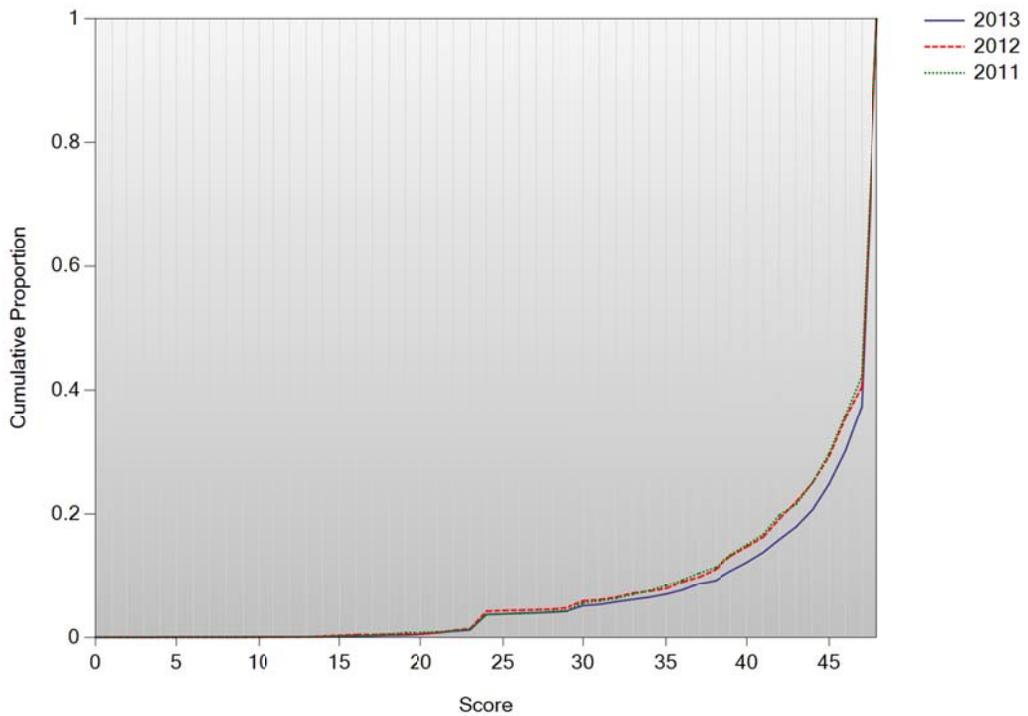
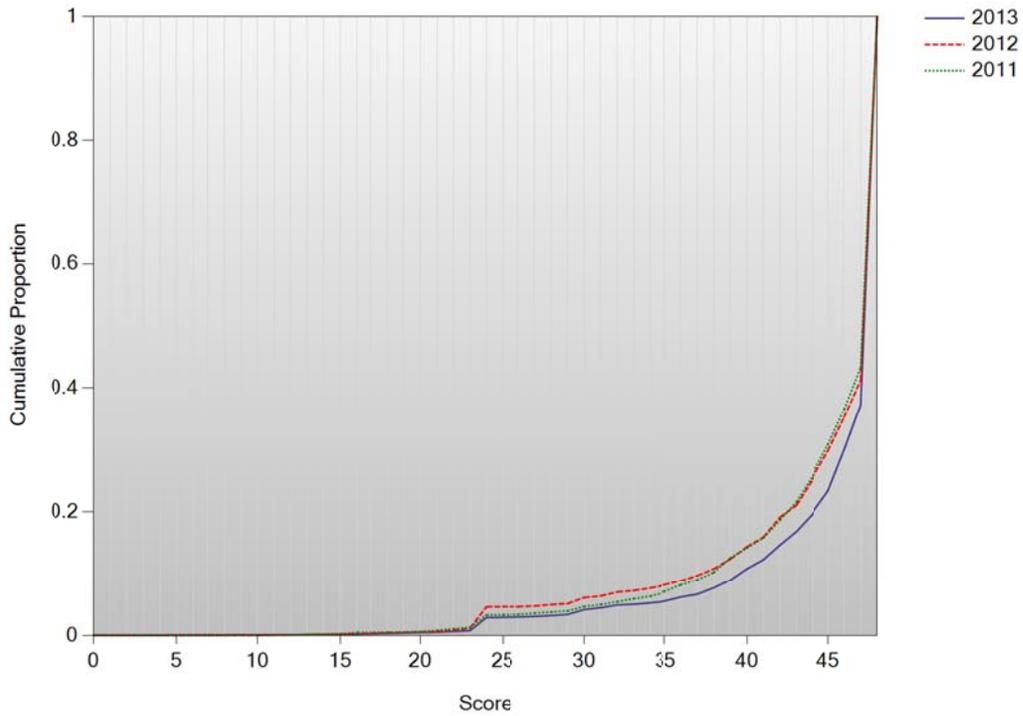


Figure G-2. 2012–13 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 5 Bottom: English Language Arts Grade 6

Cumulative Score Distributions: English Language Arts Grade 5



Cumulative Score Distributions: English Language Arts Grade 6

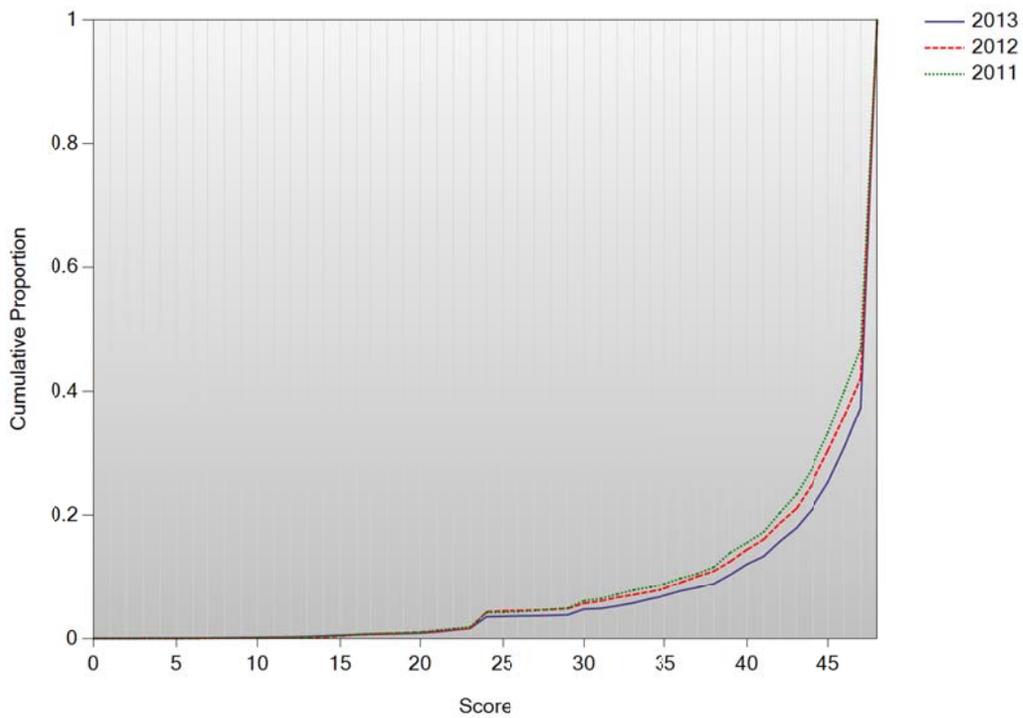
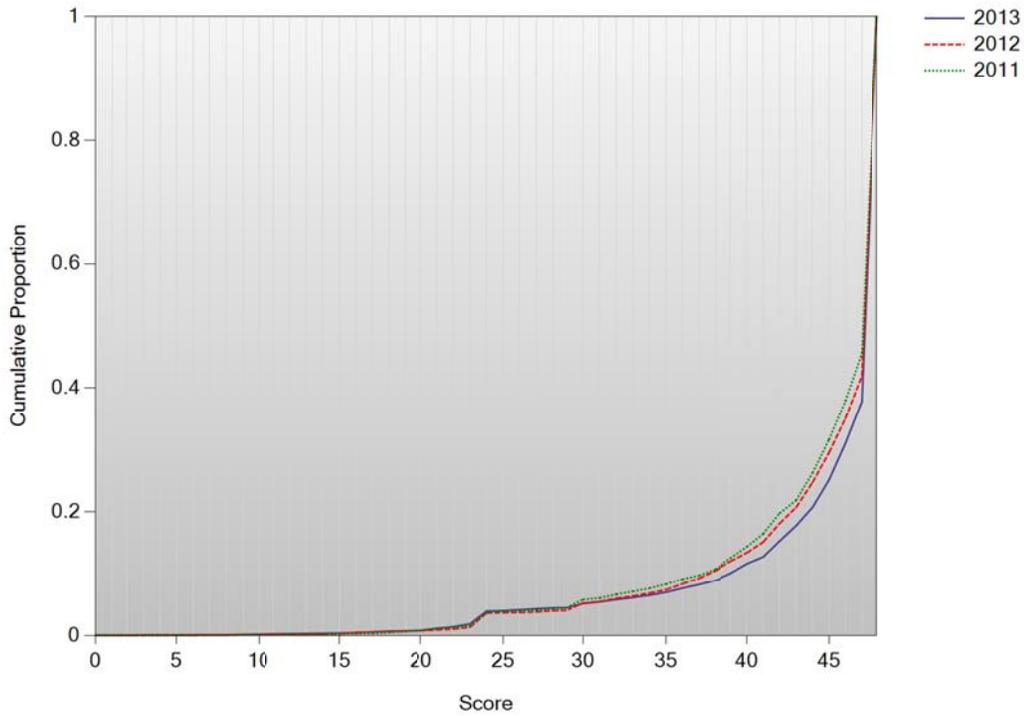


Figure G-3. 2012–13 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 7 Bottom: English Language Arts Grade 8

Cumulative Score Distributions: English Language Arts Grade 7



Cumulative Score Distributions: English Language Arts Grade 8

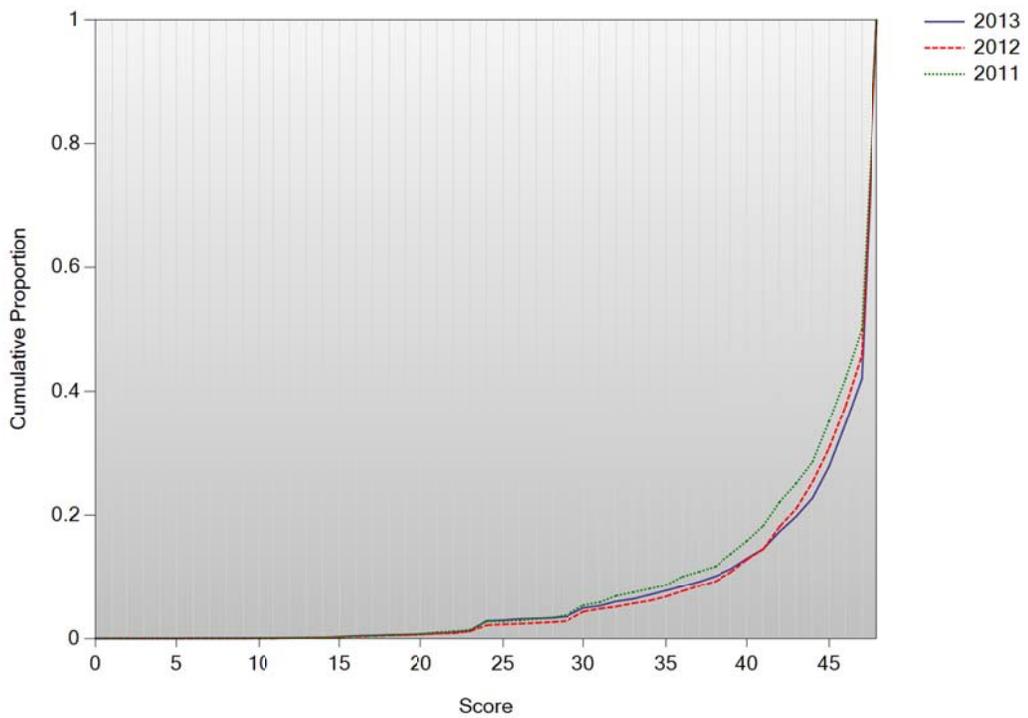
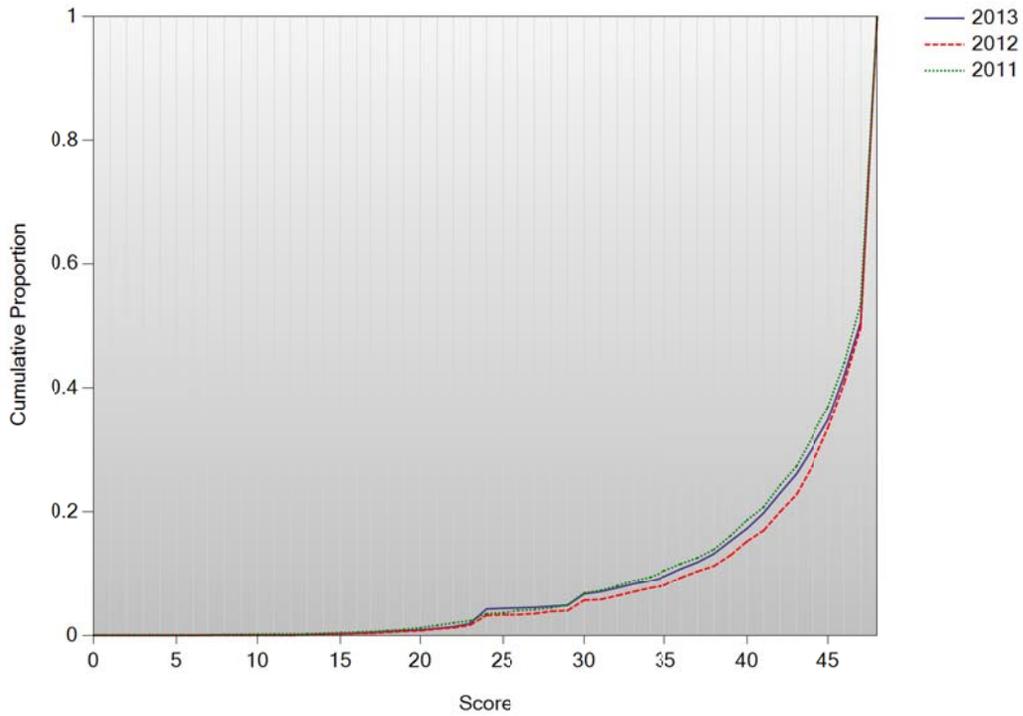


Figure G-4. 2012–13 NYSAA: Cumulative Score Distributions
Top: English Language Arts Grade 11 Bottom: Mathematics Grade 3

Cumulative Score Distributions: English Language Arts Grade 11



Cumulative Score Distributions: Mathematics Grade 3

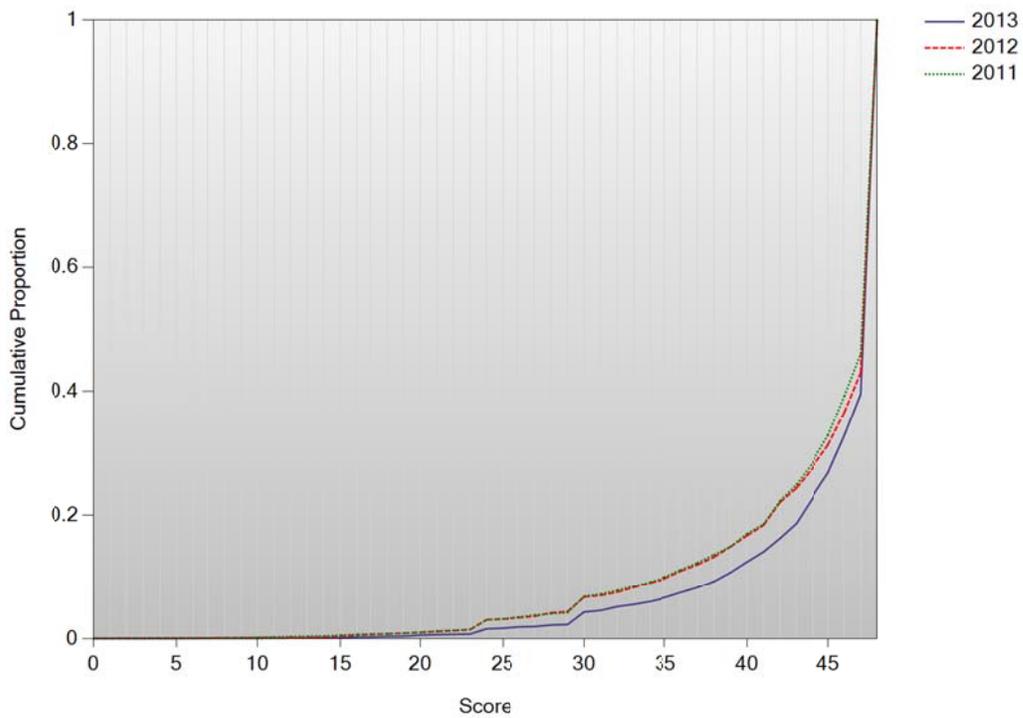
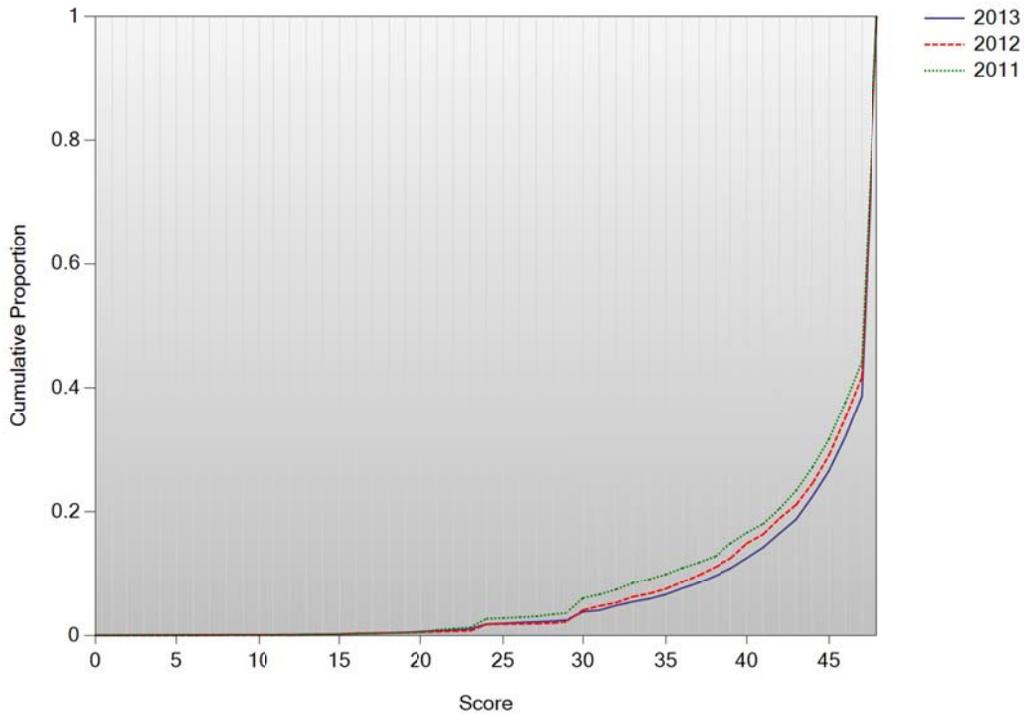


Figure G-5. 2012–13 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 4 **Bottom: Mathematics Grade 5**

Cumulative Score Distributions: Mathematics Grade 4



Cumulative Score Distributions: Mathematics Grade 5

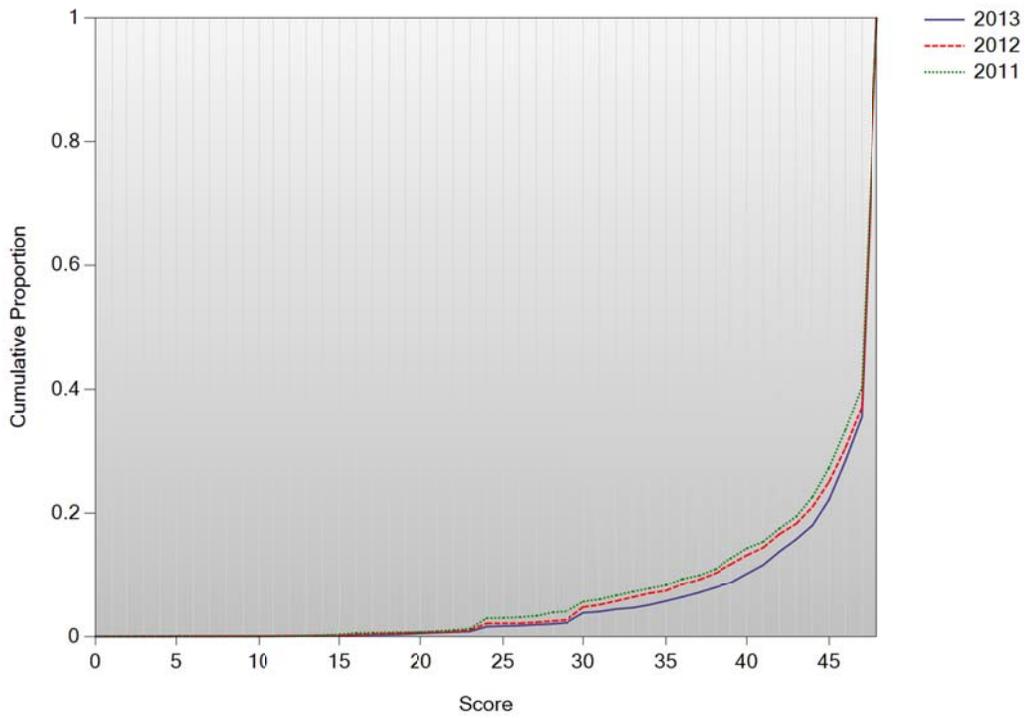
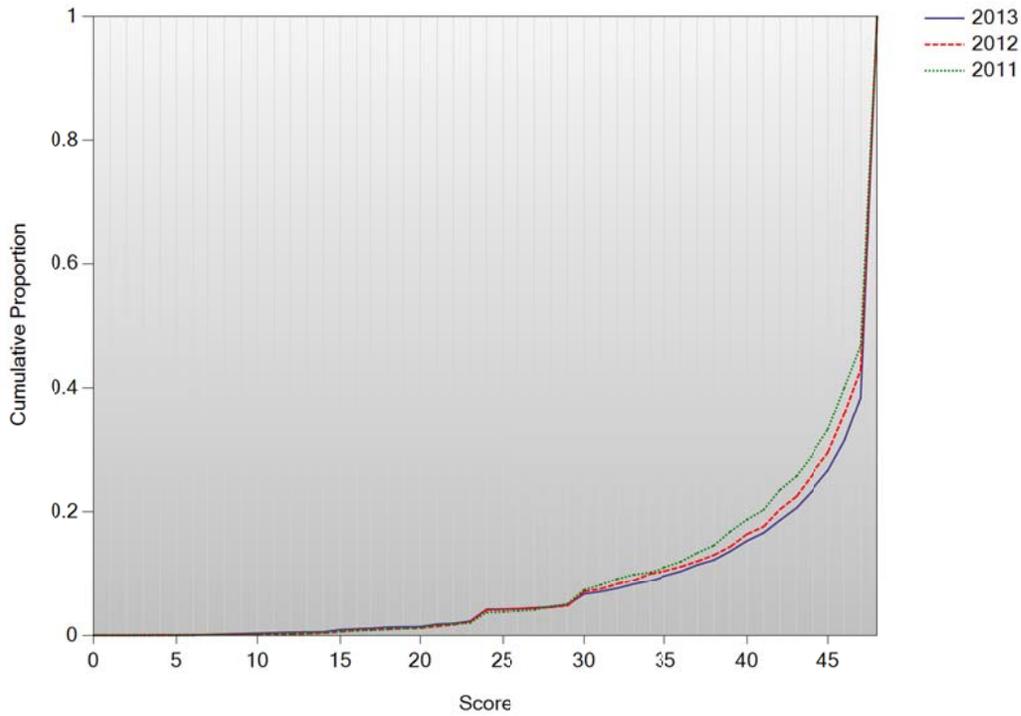


Figure G-6. 2012–13 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 6 **Bottom: Mathematics Grade 7**

Cumulative Score Distributions: Mathematics Grade 6



Cumulative Score Distributions: Mathematics Grade 7

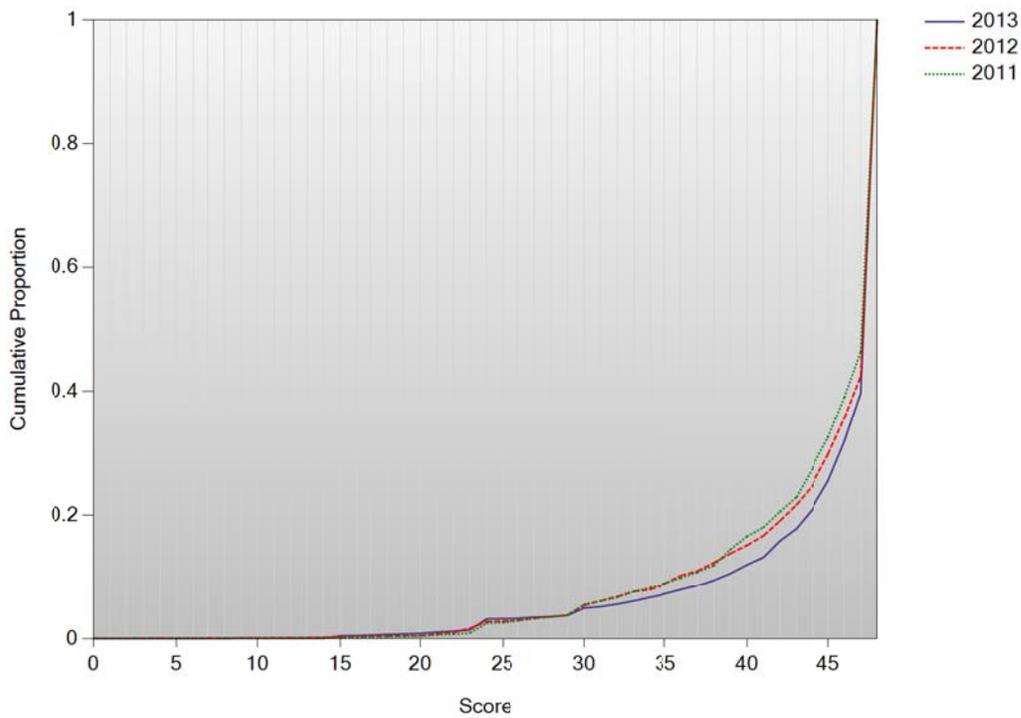
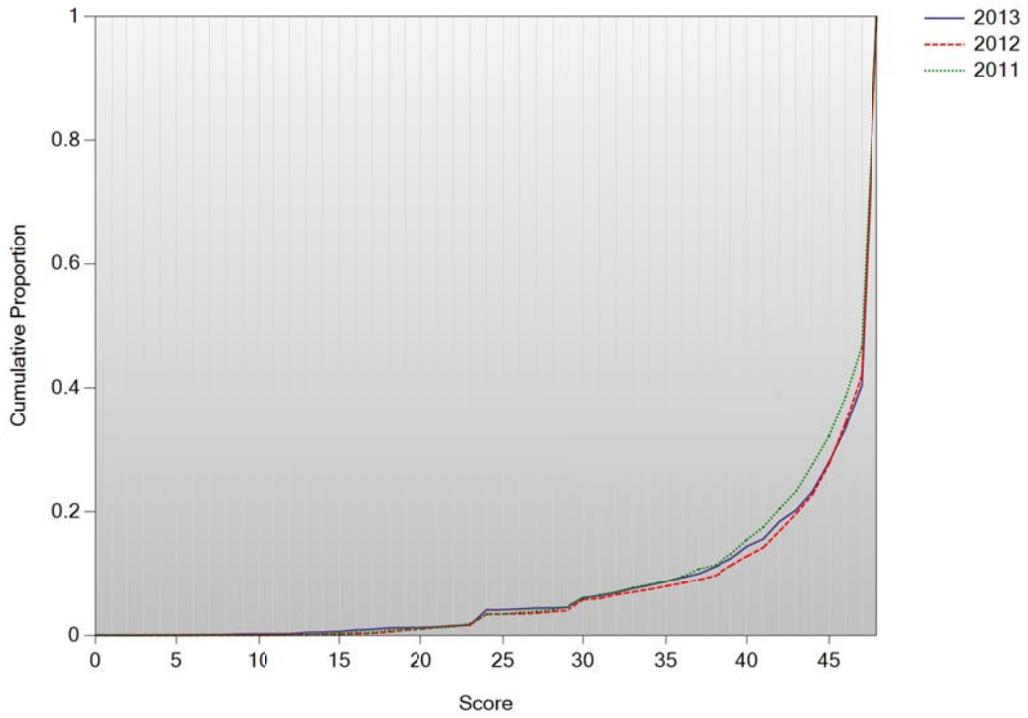


Figure G-7. 2012–13 NYSAA: Cumulative Score Distributions
Top: Mathematics Grade 8 **Bottom: Mathematics Grade 11**

Cumulative Score Distributions: Mathematics Grade 8



Cumulative Score Distributions: Mathematics Grade 11

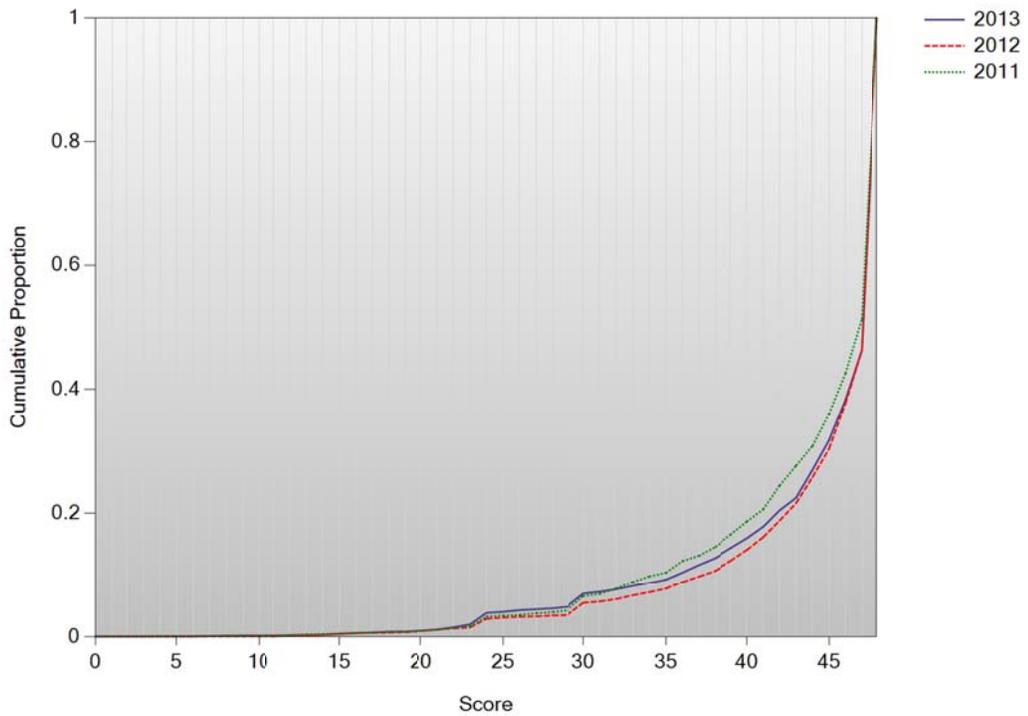
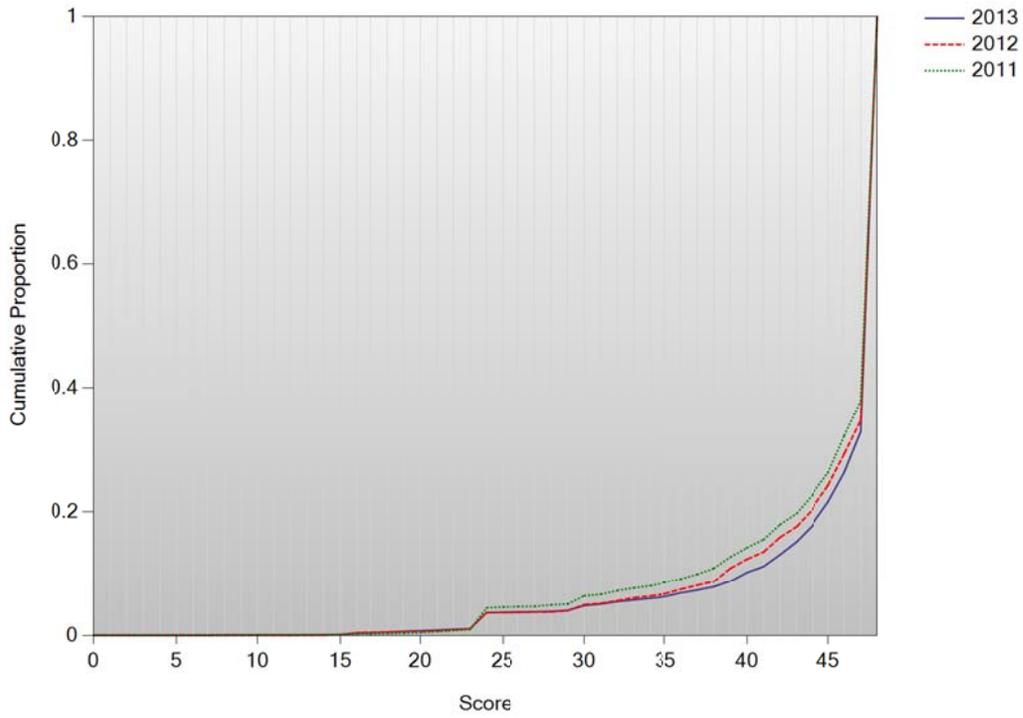


Figure G-8. 2012–13 NYSAA: Cumulative Score Distributions
Top: Science Grade 4 **Bottom: Science Grade 8**

Cumulative Score Distributions: Science Grade 4



Cumulative Score Distributions: Science Grade 8

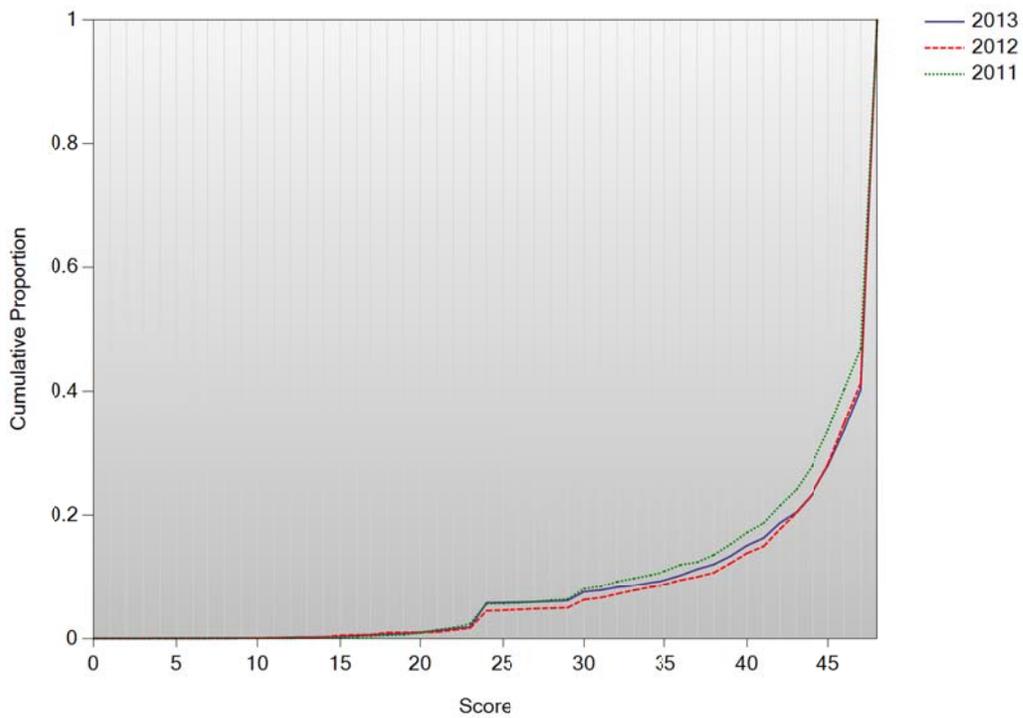
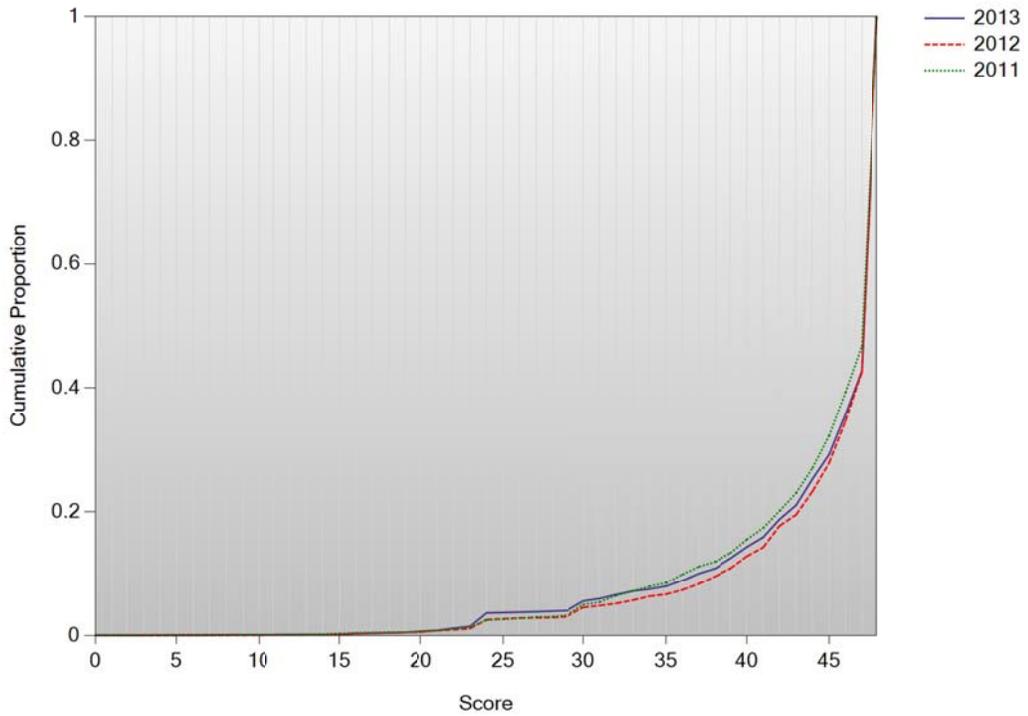


Figure G-9. 2012–13 NYSAA: Cumulative Score Distributions
Top: Science High School Bottom: Social Studies High School

Cumulative Score Distributions: Science Grade 11



Cumulative Score Distributions: Social Studies Grade 11

