

**New York State
Alternate Assessment
(NYSAA)
Standard Setting Report
June 11–13, 2014
Albany, New York**

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	DESCRIPTION OF STANDARD-SETTING METHODOLOGY.....	1
1.1	OVERVIEW OF STANDARD-SETTING PROCEDURES.....	1
1.2	ORGANIZATION OF THIS REPORT.....	2
CHAPTER 2	TASKS COMPLETED PRIOR TO THE STANDARD-SETTING MEETING	3
2.1	CREATION OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS.....	3
2.2	PREPARATION OF MATERIALS FOR PANELISTS.....	3
2.3	PREPARATION OF PRESENTATION MATERIALS	4
2.4	PREPARATION OF INSTRUCTIONS FOR FACILITATORS	4
2.5	PREPARATION OF SYSTEMS AND MATERIALS FOR ANALYSIS DURING THE MEETING	4
2.6	SELECTION OF PANELISTS.....	4
CHAPTER 3	TASKS COMPLETED DURING THE STANDARD-SETTING MEETING.....	7
3.1	OVERVIEW OF BODY OF WORK METHOD	7
3.2	ORIENTATION.....	7
3.3	REVIEW OF ASSESSMENT MATERIALS	8
3.4	REVIEW OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS	8
3.5	TRAINING ROUND.....	8
3.6	ROUND 1 JUDGMENTS.....	9
3.7	TABULATION OF ROUND 1 RESULTS.....	10
3.8	ROUND 2 JUDGMENTS.....	12
3.9	TABULATION OF ROUND 2 RESULTS.....	13
3.10	ROUND 3 JUDGMENTS.....	15
3.11	TABULATION OF ROUND 3 RESULTS.....	16
3.12	SEQUENCE OF GRADE LEVELS.....	18
3.13	EVALUATION	18
CHAPTER 4	TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING	19
4.1	CROSS-GRADE ARTICULATION COMMITTEE MEETING	19
4.2	ANALYSIS AND REVIEW OF PANELISTS' FEEDBACK.....	19
4.3	POLICY ADJUSTMENTS	20
4.4	PREPARATION OF STANDARD SETTING REPORT	20
REFERENCES	21
APPENDICES	22
Appendix A	Agenda	
Appendix B	Alternate Performance Level Descriptors	
Appendix C	Sample Student Datafolio	
Appendix D	Sample Rating Form	
Appendix E	Non-Disclosure Agreement Form	
Appendix F	Standard-Setting Process Slide Presentation	
Appendix G	Facilitator Script	

Appendix H	Panelist Summary
Appendix I	Evaluations
Appendix J	Cross-Grade Articulation Evaluation
Appendix K	Final Cuts

CHAPTER 1 DESCRIPTION OF STANDARD-SETTING METHODOLOGY

1.1 OVERVIEW OF STANDARD-SETTING PROCEDURES

The purpose of this report is to summarize the activities of the standard-setting meeting for the New York State Alternate Assessment (NYSAA) in English Language Arts (ELA) and mathematics (Grades 3–8 and High School), science (Grades 4, 8, and High School), and social studies (High School). The NYSAA standard-setting meeting was held on June 11 through 13, 2014. In all, there were 11 panels with 102 panelists participating in the process. The configuration of the panels is shown in Table 1-1.

Table 1-1. 2014 NYSAA Standard Setting: Configuration of Standard-Setting Panels

<i>Panel</i>	<i>Number of Panelists</i>	<i>Content Area(s)</i>	<i>Grade/Grade Span</i>	<i>Days</i>
1	10	ELA	3 4	Wed-Thurs Thurs
2	9	ELA	5 6	Wed-Thurs Thurs
3	9	ELA	7 8	Wed-Thurs Thurs
4	8	ELA	High School	Wed-Thurs
5	10	Math	3 4	Wed-Thurs Thurs
6	9	Math	5 6	Wed-Thurs Thurs
7	9	Math	7 8	Wed-Thurs Thurs
8	9	Math	High School	Wed-Thurs
9	10	Science	4 8	Wed-Thurs Thurs
10	9	Science	High School	Wed-Thurs
11	10	Social Studies	High School	Wed-Thurs

The body of work method was used for setting standards for the NYSAA. In the body of work method, panelists are presented with samples of actual student work (in this case, student datafolios) and make their judgments based on those samples. Specifically, panelists examine each student datafolio and, based on their common understanding of the Alternate Performance Level Descriptors (APLDs), determine which performance level best matches the evidence that the student exhibits through his or her performance on the assessment. The agenda for the standard-setting meeting is provided in Appendix A.

1.2 ORGANIZATION OF THIS REPORT

This report is organized into three major sections, describing tasks completed prior to (1), during (2), and after (3) the standard-setting meeting.

CHAPTER 2 TASKS COMPLETED PRIOR TO THE STANDARD-SETTING MEETING

2.1 CREATION OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS

The Alternate Performance Level Descriptors (APLDs) for each grade and content area provided panelists with the official description of the knowledge, skills, and understanding that students are expected to be able to display to be classified into each performance level. The APLDs were developed using the old APLDs and the general education Performance Level Descriptors for Grades 3 through 8 English Language Arts (ELA) and mathematics. The initial language was developed by the Regional Lead Trainers and was then refined by Measured Progress. The APLDs were reviewed, edited, and approved by the New York State Education Department (the Department). The APLDs utilized during standard setting are included in Appendix B.

2.2 PREPARATION OF MATERIALS FOR PANELISTS

The following materials were assembled for presentation to the panelists at the standard-setting meeting:

- Meeting agenda
- Non-Disclosure Agreement form
- APLDs
- Student datafolios—sets of datafolios of actual student work, ordered from lowest to highest score
- Rating sheets containing datafolio identifiers and four columns where the panelists marked their ratings—each column corresponded to the four performance levels (Not Meeting Learning Standards, Partially Meeting Learning Standards, Meeting Learning Standards, Meeting Learning Standards with Distinction)
- Evaluation forms

Additionally, the following materials were available to the panelists at the standard-setting meeting:

- 2013-14 New York State Alternate Assessment (NYSAA) Administration Manual
- Sample student datafolio (Appendix B of the NYSAA Administration Manual)
- 2013-14 NYSAA Frameworks (Appendix F of the NYSAA Administration Manual)

Copies of the meeting agenda, the APLDs, the Non-Disclosure Agreement form, the sample student datafolio, and the rating form are included in Appendices A through E.

2.3 PREPARATION OF PRESENTATION MATERIALS

The Standard-Setting Process slide presentation used in the opening session was prepared prior to the meeting. The slide presentation was reviewed, edited, and approved by the Department. A copy of the presentation is included in Appendix F.

2.4 PREPARATION OF INSTRUCTIONS FOR FACILITATORS

A script was created for the group facilitators to refer to while working through each step of the standard-setting process. The script was provided to the Department for their records. This document is included in Appendix G. The facilitators also attended a training session led by a Measured Progress psychometrician and NYSAA program management two weeks before the standard-setting meeting. The purpose of the training was to prepare the facilitators for the panel activities and to ensure consistency in the implementation of procedures. The facilitators also attended a meeting with the Department prior to the standard-setting meeting. The purpose of the meeting was to allow the Department to share information about the NYSAA and to allow for any questions that the facilitators may have had. This ensured that the facilitators had the same foundation regarding the NYSAA.

2.5 PREPARATION OF SYSTEMS AND MATERIALS FOR ANALYSIS DURING THE MEETING

The computational programming used to calculate cut scores and impact data during the standard-setting meeting was completed and thoroughly tested prior to the standard-setting meeting. All calculations and formats for data used as inputs to standard-setting programs were parallel processed (for example, applying student-level exclusions); all programs used to calculate impact data were parallel processed; and cut score programs were carefully reviewed and thoroughly tested by conducting a code review and entering in mock panelist ratings and reviewing results. See Section 3.7, Tabulation of Round 1 Results, for a description of the analyses performed during standard setting.

2.6 SELECTION OF PANELISTS

Panelists were recruited and selected by the Department prior to the standard-setting session. The panels were intended to be made up of 10 panelists each, with a total of 110 participants. Because there were many stakeholders with great and sincere interest in the outcomes of standard setting, it was important that they be well represented. Each panel was composed of a diverse group of stakeholders, including general education and special education teachers, school administrators, higher education personnel, and/or stakeholders from disability advocate groups. Additionally, the selection of panelists reflected a balance of gender, race/ethnicity, and geographic location. The selection criteria for panelists also included that they be familiar with both the subject matter and/or the student population

assessed via the NYSAA, as well as the grade(s) for which they would be setting standards and recommending performance standards.

Panelists were recruited through various groups and organizations. The Department utilized an online application process. The NYSAA standard-setting application was posted to the Department's Web site in October 2013 and was available online through February 2014. The following list includes several of the groups and organizations that were contacted for panel participation:

- New York State Alternate Assessment Training Network (AATN) Specialists
- New York State Alternate Assessment Score Site Coordinators (SSCs)
- New York State Alternate Assessment Alignment Work Group
- Big Five City School Districts in New York State
- Regional Information Centers in New York State
- Assessment Directors in New York State
- Committees on Special Education (CSEs) throughout New York State
- Boards of Cooperative Educational Services (BOCES)
- Superintendents of State-Operated Schools
- New York State United Teachers (NYSUT)
- General and Special Education Teachers (open application process) throughout New York State

The actual number of panelists who participated was 102, distributed into 11 panels, as shown in Table 2-1, with demographic information presented in Appendix H. All panelists provided voluntary demographic information. The demographic information summarized came from the application and/or the evaluation. Table 2-2 presents a summary of all panelists by gender representation. Table 2-3 presents a summary of all panelists by ethnic representation. Table 2-4 presents a summary of all panelists by geographic locations. Table 2-5 presents a summary of the current positions of all panelists. Also important to note is that 56 of the 102 panelists indicated that they were currently administering the NYSAA.

Table 2-1. 2014 NYSAA Standard Setting: Configuration of Standard-Setting Panels

Content	Grade 3/4	Grade 5/6	Grade 7/8	High School
ELA	10	9	9	8
Mathematics	10	9	9	9
Science	10 (Grades 4 and 8)			9
Social Studies				10

Table 2-2. 2014 NYSAA Standard Setting: Panelists by Gender

Title	Number
Female	88
Male	14

Table 2-3. 2014 NYSAA Standard Setting: Panelists by Ethnicity

Title	Number
Asian	2
Black/African American	19
Hispanic/Latino	3
White	76
Unassigned	2

Table 2-4. 2014 NYSAA Standard Setting: Panelists by Geographic Region/Big City

Title	Number
Big City (not including NYC)	13
Capital Region	7
Central NY	11
Long Island	9
Lower and Mid-Hudson Valley	13
North County	2
NYC	31
Western NY	10
Southern Tier	6

Table 2-5. 2014 NYSAA Standard Setting: Panelists by Current Position*

Title	Number
Special Education Teacher	50
General Education Content Teacher	24
Administrator/AATN Specialist/SSC	19
Institutes of Higher Education	6
Other**	3
Total	102

*Please note: panelists were able to indicate that they held more than one current position.

**Some responses in the Other category include United Federation of Teachers (UFT) Teacher Center, School Psychologist, English Language Learner (ELL) Teacher, and Professional Development Specialist.

CHAPTER 3 TASKS COMPLETED DURING THE STANDARD-SETTING MEETING

3.1 OVERVIEW OF BODY OF WORK METHOD

The body of work standard-setting method was developed specifically for use with assessments that are designed to allow for a range of student responses, such as portfolio- and performance-based assessments (Kingston et al., 2001). Panelists were asked to classify each student datafolio into a single performance level by considering the evidence that the student provided in the datafolio. Scores for the datafolios were not provided to the panelists.

3.2 ORIENTATION

With regard to panelist training, *Standards for Educational and Psychological Testing* states the following:

Care must be taken to assure these persons understand what they are to do and that their judgments are as thoughtful and objective as possible. The process must be such that well-qualified participants can apply their knowledge and experience to reach meaningful and relevant judgments that accurately reflect their understandings and intentions.

(AERA/APA/NCME, 2014, p. 101)

The training of the panelists began with a general orientation at the start of the standard-setting meeting. The purpose of the orientation was to ensure that all panelists received the same information about the need for and goals of standard setting and about their part in the process. First, a representative from the New York State Education Department (the Department) provided some pertinent context about the New York State Alternate Assessment (NYSAA) redesign process and an introduction to the issues of standard setting. Included in this introduction was information regarding setting constraints, or guardrails, on the cut score between Partially Meeting Learning Standards and Meeting Learning Standards, so that at least 70% of the students were categorized as Meeting Learning Standards or Meeting Learning Standards with Distinction. The rationale for this decision was based on the 30% decrease in the number of students scoring at the proficient level in the 3-8 general assessments and the NYSAA cut scores needed to reflect this decrease. Second, Measured Progress Special Education personnel provided an overview of the assessment, including administration, scoring, and video clips of a variety of students during administration of the alternate assessment. Next, a Measured Progress psychometrician presented an overview of the body of work procedure and the activities that would occur during the standard-setting meeting sessions. Once the general orientation

was complete, each panel convened in a break-out room, where the panelists received more detailed training from their facilitator and completed the standard-setting activities.

3.3 REVIEW OF ASSESSMENT MATERIALS

The first step after the opening session was for the panelists to become familiar with the NYSAA. The purpose of this step was to make sure that the panelists thoroughly understood the NYSAA Frameworks and how the assessment is administered and scored. Panelists reviewed the quick reference guide, the practice datafolio sets, the scoring rubric, and the NYSAA Frameworks. Panelists engaged in a discussion of the administration of the assessment, the student participants, the scoring rubric, and other pertinent information. The 2013-14 NYSAA Administration Manual, including a sample datafolio in Appendix B of the manual, was also provided to each group as a reference document.

3.4 REVIEW OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS

The second step in the process was to discuss the Alternate Performance Level Descriptors (APLDs). This important step was designed to ensure that panelists thoroughly understood the knowledge, skills, and understanding needed for students to be classified into performance levels (Not Meeting Learning Standards, Partially Meeting Learning Standards, Meeting Learning Standards, Meeting Learning Standards with Distinction). Panelists first reviewed the APLDs on their own and then participated in group discussions of the APLDs, clarifying the description for each performance level. The discussions focused on the evidence that differentiated adjacent performance levels. Bulleted lists of characteristics for each level were generated based on the group discussion, and were posted in the room for panelists to refer to during the rounds of ratings.

Note that the purpose of this step was to clarify and add specificity to the APLDs based on the knowledge, skills, and understanding, as well as the combination of the performance that the evidence showed in relation to the APLDs identified for each practice datafolio in the previous step (Review of Assessment Materials). The bulleted lists were developed as working documents to be used by the panelists for the purposes of standard setting. They supplemented the APLDs, which provide the official definition of what it means for a student to be classified into each performance level, by specifically addressing the knowledge, skills, and understanding measured by the NYSAA. The APLDs are provided in Appendix B.

3.5 TRAINING ROUND

Next, the panelists completed a practice round of ratings. The purpose of the practice round was to familiarize the panelists with all of the materials they would be using as part of the standard-setting

process and to walk them through the process of rating student datafolios. In addition to the APLDs, panelists were given the following materials:

Practice student datafolios and rating form. The panelists were given a set of three practice student datafolios. They were also given a rating form, which provided a column for the datafolio identifier and four blank columns. The blank columns represented the four performance levels (Not Meeting Learning Standards, Partially Meeting Learning Standards, Meeting Learning Standards, Meeting Learning Standards with Distinction). The panelists entered a checkmark in the column that they believed that the student datafolio should be classified in. A sample rating form is provided in Appendix D; the practice rating form had the same structure, except that only three datafolios were included.

The facilitator explained each aspect of the assessment material and how panelists would use it to make their ratings. Then the facilitator reviewed the first datafolio with the panelists, pointing out the evidence contained in the datafolio. The facilitator reviewed the relationship between the evidence provided by the datafolio and the relationship to the APLDs. The second and third datafolios were reviewed with panelists in the same manner. Panelists were asked to rate each datafolio, focusing on the clarifications agreed upon by the group in relation to the APLDs. The facilitator then led the panelists through a readiness discussion, asking them to share their reasoning for each datafolio rating and assessing each panelist's understanding of the rating task.

3.6 ROUND 1 JUDGMENTS

In the first round, panelists worked individually with the APLDs and the student datafolios and rating form. The rating form consisted of 25 to 30 datafolios, with scores covering the range of obtained scores. For each datafolio, the panelists considered the evidence and the knowledge, skills, and understanding demonstrated by the student in that datafolio, and panelists decided which performance level was the best match. The panelists worked their way through the datafolios, making a rating for each one, and recorded their ratings in the Round 1 rating form. While the datafolios were presented in order of total score, panelists were not required to rate them strictly in increasing order. Instead, panelists were encouraged to take a holistic look at the evidence in the datafolio to make a judgment about the appropriate performance level relative to the clarifications agreed upon by the group in relation to the APLDs.

In one content area/grade level combination, English Language Arts (ELA) High School, there was an anomalous result after Round 1 because there was no cut score for the top category. Essentially, the cut score was placed so high that it was outside of the score scale range. Upon review of the data from the panelists and discussions with the facilitator, it was determined that the panelists were not clear on their definition of Meeting Learning Standards with Distinction, making classifications between Meeting Learning Standards and Meeting Learning Standards with Distinction unclear. To

obtain cut scores (and impact data) for this grade/content area, it was necessary to remove some ratings. Once the ratings were removed from the results, a cut score resulted on the score scale range. These are the results presented in the Round 1 Table (Table 3-1).

3.7 TABULATION OF ROUND 1 RESULTS

After all panelists had completed their individual ratings, the Measured Progress data analysis team calculated the average cut scores for the group based on the Round 1 ratings. Cut scores were calculated using Statistical Analysis Software (SAS) statistical software. Logistic regression was used to determine each panelist’s individual cut scores, and then the cut scores were averaged across the group. In addition, the standard error of the panelists’ cut scores was calculated, which provided an indication of the extent to which judgments were consistent across panelists and, in particular, reflected the increasing level of agreement between the ratings with each successive round of ratings. The Round 1 results are outlined in Tables 3-1 through 3-4.

Table 3-1. 2014 NYSAA Standard Setting: Round 1 Results—ELA

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	126	3
	Partially Meeting Learning Standards	126.9	2.6	127	156	38
	Meeting Learning Standards	156.3	5.2	157	190	52
	Meeting Learning Standards with Distinction	190.3	6.5	191	225	7
4	Not Meeting Learning Standards	NA	NA	100	125	3
	Partially Meeting Learning Standards	125.9	1.5	126	150	32
	Meeting Learning Standards	150.2	3.6	151	188	58
	Meeting Learning Standards with Distinction	188.9	3.4	189	225	7
5	Not Meeting Learning Standards	NA	NA	100	134	9
	Partially Meeting Learning Standards	134.8	1.3	135	149	23
	Meeting Learning Standards	149.9	3.1	150	217	68
	Meeting Learning Standards with Distinction	217.2	6.9	218	225	0
6	Not Meeting Learning Standards	NA	NA	100	130	5
	Partially Meeting Learning Standards	130.6	1.4	131	148	18
	Meeting Learning Standards	148.5	1.6	149	188	72
	Meeting Learning Standards with Distinction	188.7	1.7	189	225	6
7	Not Meeting Learning Standards	NA	NA	100	129	6
	Partially Meeting Learning Standards	129.7	2.8	130	155	57
	Meeting Learning Standards	155.9	9.8	156	179	27
	Meeting Learning Standards with Distinction	179.6	12.1	180	225	10
8	Not Meeting Learning Standards	NA	NA	100	137	15
	Partially Meeting Learning Standards	137.6	1.8	138	167	70
	Meeting Learning Standards	167.2	2.3	168	197	12
	Meeting Learning Standards with Distinction	198.0	2.9	198	225	3
HS	Not Meeting Learning Standards	NA	NA	100	123	3
	Partially Meeting Learning Standards	123.4	36.8	124	156	39
	Meeting Learning Standards	156.6	6.9	157	208	58
	Meeting Learning Standards with Distinction	208.1	4.8	209	225	1

Table 3-2. 2014 NYSAA Standard Setting: Round 1 Results—Mathematics

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	136	8
	Partially Meeting Learning Standards	136.5	2.7	137	183	81
	Meeting Learning Standards	183.2	4.7	184	204	9
	Meeting Learning Standards with Distinction	204.1	3.1	205	225	2
4	Not Meeting Learning Standards	NA	NA	100	134	11
	Partially Meeting Learning Standards	134.7	2.4	135	167	61
	Meeting Learning Standards	167.1	2.9	168	199	25
	Meeting Learning Standards with Distinction	199.7	3.0	200	225	4
5	Not Meeting Learning Standards	NA	NA	100	124	5
	Partially Meeting Learning Standards	124.3	0.7	125	149	35
	Meeting Learning Standards	149.6	0.7	150	190	56
	Meeting Learning Standards with Distinction	190.6	0.7	191	225	4
6	Not Meeting Learning Standards	NA	NA	100	128	6
	Partially Meeting Learning Standards	129.0	0.0	129	147	24
	Meeting Learning Standards	148.0	0.0	148	183	59
	Meeting Learning Standards with Distinction	184.0	0.0	184	225	11
7	Not Meeting Learning Standards	NA	NA	100	125	3
	Partially Meeting Learning Standards	125.4	0.4	126	148	24
	Meeting Learning Standards	148.7	2.6	149	185	61
	Meeting Learning Standards with Distinction	186.0	3.7	186	225	11
8	Not Meeting Learning Standards	NA	NA	100	132	10
	Partially Meeting Learning Standards	133.0	1.2	133	158	58
	Meeting Learning Standards	158.8	2.6	159	205	31
	Meeting Learning Standards with Distinction	205.9	2.8	206	225	2
HS	Not Meeting Learning Standards	NA	NA	100	127	5
	Partially Meeting Learning Standards	128.0	0.8	128	142	10
	Meeting Learning Standards	143.0	2.0	143	208	84
	Meeting Learning Standards with Distinction	209.0	7.1	209	225	1

Table 3-3. 2014 NYSAA Standard Setting: Round 1 Results—Science

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
4	Not Meeting Learning Standards	NA	NA	100	107	2
	Partially Meeting Learning Standards	107.7	0.6	108	117	19
	Meeting Learning Standards	117.5	4.0	118	128	59
	Meeting Learning Standards with Distinction	128.9	17.0	129	150	21
8	Not Meeting Learning Standards	NA	NA	100	111	5
	Partially Meeting Learning Standards	111.9	0.5	112	121	31
	Meeting Learning Standards	121.1	1.1	122	146	61
	Meeting Learning Standards with Distinction	146.7	1.6	147	150	3
HS	Not Meeting Learning Standards	NA	NA	100	108	2
	Partially Meeting Learning Standards	108.9	1.1	109	124	38
	Meeting Learning Standards	124.2	1.2	125	146	57
	Meeting Learning Standards with Distinction	146.6	1.9	147	150	3

Table 3-4. 2014 NYSAA Standard Setting: Round 1 Results—Social Studies

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
HS	Not Meeting Learning Standards	NA	NA	100	109	3
	Partially Meeting Learning Standards	110.0	1.2	110	120	25
	Meeting Learning Standards	120.4	1.9	121	136	64
	Meeting Learning Standards with Distinction	136.7	2.0	137	150	7

3.8 ROUND 2 JUDGMENTS

The purpose of Round 2 was for panelists to discuss their Round 1 judgments as a group and determine whether any revisions to the clarifications agreed upon by the group in relation to the APLDs were necessary. A psychometrician shared the average cut score locations and impact data (i.e., the percentage of students classified into each performance level based on the group average cuts) with the panelists to help inform their group discussion and Round 2. The psychometrician explained how to use the information as they completed their Round 2 discussions. Panelists were encouraged to discuss whether the percentage of students classified in each performance level seemed reasonable, given their perceptions of the students and the evidence required to show the students' levels of performance. Impact data are presented in Tables 3-1 through 3-4. Prior to the group discussion, the facilitator asked for a show of hands to determine the number of panelists who had placed each datafolio into each performance level; the facilitator then recorded the results on chart paper. Starting with the first datafolio they disagreed on, the panelists began discussing the categorization of the datafolios according to their initial ratings in the context of the classifications made by other members of the group. Panelists were encouraged to share their own points of view as well as to listen to the comments of their colleagues. Facilitators made sure that the panelists knew that the purpose of the discussion was not to reach consensus; at every point throughout the standard-setting process, the panelists were asked to provide

his or her own best judgment. Once the discussions were complete, the panelists filled in the Round 2 column of their rating form.

In each of two content area/grade level combinations, English Language Arts (ELA) grade 5 and Science High School, there was an anomalous result after Round 2 because there was no cut score for the top category. Essentially, the cut score was placed so high that it was outside of the score scale range. Upon review of the data from the panelists and discussions with the facilitator, it was determined that the panelists were not clear on their definition of Meeting Learning Standards with Distinction, making classifications between Meeting Learning Standards and Meeting Learning Standards with Distinction unclear. To obtain cut scores (and impact data) for this grade/content area, it was necessary to remove some ratings. Once the ratings were removed from the results, a cut score resulted on the score scale. These are the results presented in the Round 2 Tables 3-5 and 3-7.

3.9 TABULATION OF ROUND 2 RESULTS

When Round 2 ratings were complete, the Measured Progress data analysis team calculated the average cut scores for the room and associated impact data. The results of the panelists' Round 2 ratings are outlined in Tables 3-5 through 3-8.

Table 3-5. 2014 NYSAA Standard Setting: Round 2 Results—ELA

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	129	4
	Partially Meeting Learning Standards	129.7	1.2	130	152	29
	Meeting Learning Standards	152.8	2.9	153	190	60
	Meeting Learning Standards with Distinction	190.4	5.7	191	225	7
4	Not Meeting Learning Standards	NA	NA	100	125	3
	Partially Meeting Learning Standards	126.0	0.8	126	147	22
	Meeting Learning Standards	148.0	2.4	148	189	68
	Meeting Learning Standards with Distinction	189.2	2.7	190	225	7
5	Not Meeting Learning Standards	NA	NA	100	135	9
	Partially Meeting Learning Standards	135.5	1.0	136	146	15
	Meeting Learning Standards	147.0	0.9	147	202	74
	Meeting Learning Standards with Distinction	203.0	3.9	203	225	1
6	Not Meeting Learning Standards	NA	NA	100	130	5
	Partially Meeting Learning Standards	130.4	1.2	131	146	14
	Meeting Learning Standards	146.6	1.5	147	185	72
	Meeting Learning Standards with Distinction	185.5	1.0	186	225	9
7	Not Meeting Learning Standards	NA	NA	100	130	8
	Partially Meeting Learning Standards	130.7	2.3	131	157	62
	Meeting Learning Standards	157.5	3.6	158	202	28
	Meeting Learning Standards with Distinction	202.9	3.8	203	225	2

continued

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
8	Not Meeting Learning Standards	NA	NA	100	134	11
	Partially Meeting Learning Standards	134.2	1.5	135	165	73
	Meeting Learning Standards	166.0	1.8	166	195	14
	Meeting Learning Standards with Distinction	195.8	0.2	196	225	3
HS	Not Meeting Learning Standards	NA	NA	100	130	5
	Partially Meeting Learning Standards	130.1	0.6	131	154	32
	Meeting Learning Standards	154.3	4.6	155	198	59
	Meeting Learning Standards with Distinction	198.2	3.9	199	225	3

Table 3-6. 2014 NYSAA Standard Setting: Round 2 Results—Mathematics

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	134	7
	Partially Meeting Learning Standards	134.6	2.4	135	176	72
	Meeting Learning Standards	176.8	1.2	177	205	19
	Meeting Learning Standards with Distinction	205.3	0.7	206	225	2
4	Not Meeting Learning Standards	NA	NA	100	135	12
	Partially Meeting Learning Standards	135.3	1.0	136	162	50
	Meeting Learning Standards	162.2	1.1	163	195	34
	Meeting Learning Standards with Distinction	195.1	1.5	196	225	5
5	Not Meeting Learning Standards	NA	NA	100	115	2
	Partially Meeting Learning Standards	115.8	0.0	116	152	53
	Meeting Learning Standards	152.6	1.2	153	189	41
	Meeting Learning Standards with Distinction	189.5	1.5	190	225	4
6	Not Meeting Learning Standards	NA	NA	100	128	6
	Partially Meeting Learning Standards	129.0	0.0	129	147	24
	Meeting Learning Standards	148.0	0.0	148	183	59
	Meeting Learning Standards with Distinction	184.0	0.0	184	225	11
7	Not Meeting Learning Standards	NA	NA	100	125	3
	Partially Meeting Learning Standards	125.1	0.7	126	157	48
	Meeting Learning Standards	157.8	1.4	158	193	42
	Meeting Learning Standards with Distinction	193.2	1.3	194	225	7
8	Not Meeting Learning Standards	NA	NA	100	132	10
	Partially Meeting Learning Standards	132.9	0.5	133	159	59
	Meeting Learning Standards	159.8	2.3	160	204	30
	Meeting Learning Standards with Distinction	204.8	2.3	205	225	2
HS	Not Meeting Learning Standards	NA	NA	100	129	6
	Partially Meeting Learning Standards	129.1	1.0	130	141	9
	Meeting Learning Standards	141.2	1.2	142	198	81
	Meeting Learning Standards with Distinction	198.3	3.8	199	225	4

Table 3-7. 2014 NYSAA Standard Setting: Round 2 Results—Science

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
4	Not Meeting Learning Standards	NA	NA	100	109	3
	Partially Meeting Learning Standards	109.3	0.4	110	117	18
	Meeting Learning Standards	117.6	0.9	118	144	78
	Meeting Learning Standards with Distinction	144.8	3.2	145	150	1
8	Not Meeting Learning Standards	NA	NA	100	112	6
	Partially Meeting Learning Standards	112.3	0.2	113	120	29
	Meeting Learning Standards	120.9	1.1	121	143	62
	Meeting Learning Standards with Distinction	143.8	0.9	144	150	3
HS	Not Meeting Learning Standards	NA	NA	100	109	3
	Partially Meeting Learning Standards	109.2	0.7	110	121	34
	Meeting Learning Standards	121.8	1.2	122	149	61
	Meeting Learning Standards with Distinction	149.4	1.4	150	150	2

Table 3-8. 2014 NYSAA Standard Setting: Round 2 Results—Social Studies

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
HS	Not Meeting Learning Standards	NA	NA	100	108	3
	Partially Meeting Learning Standards	108.6	0.7	109	119	15
	Meeting Learning Standards	119.7	1.5	120	134	48
	Meeting Learning Standards with Distinction	134.3	1.4	135	150	35

3.10 ROUND 3 JUDGMENTS

The purpose of Round 3 was for panelists to discuss their Round 2 ratings as a whole group and, if necessary, to revise their judgments. Prior to the group discussion, the facilitator once again asked for a show of hands to determine the number of panelists who had placed each datafolio into each performance level; the facilitator recorded the results on chart paper. The group average cuts based on the Round 2 results were presented. In addition, in this round, the group was again presented with the impact data (i.e., the percentage of students classified into each performance level based on the group average cuts), as well as the historical Performance data (based on the previous year's assessment). The psychometrician presented the group average cuts and both sets of impact data to the group and explained how to use the information as they completed their Round 3 discussions. Panelists were encouraged to discuss whether the percentage of students classified in each performance level seemed reasonable, given their perceptions of the students and the evidence required to show the students' levels of performance. In general, panelists were not surprised by the change in the performance data, given the redesign of the assessment. As in Round 2, starting with the first datafolio for which there was disagreement, the panelists discussed their ratings, with the impact data considered as additional context for the discussion. Finally, after the discussions were complete,

panelists were given a final opportunity to revise their ratings. Once again, the facilitator reminded the panelists that they should use their individual best judgment and that it was not necessary for them to reach a consensus.

3.11 TABULATION OF ROUND 3 RESULTS

When Round 3 ratings were complete, the Measured Progress data analysis team once again calculated the average cut scores for the room and the associated impact data. The results of the panelists' Round 3 ratings are outlined in Tables 3-9 through 3-12.

Table 3-9. 2014 NYSAA Standard Setting: Round 3 Results—ELA

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	130	5
	Partially Meeting Learning Standards	130.4	1.0	131	154	31
	Meeting Learning Standards	154.2	2.2	155	192	57
	Meeting Learning Standards with Distinction	192.7	4.3	193	225	6
4	Not Meeting Learning Standards	NA	NA	100	128	4
	Partially Meeting Learning Standards	128.4	0.7	129	146	19
	Meeting Learning Standards	146.2	2.2	147	190	71
	Meeting Learning Standards with Distinction	190.3	2.5	191	225	6
5	Not Meeting Learning Standards	NA	NA	100	135	9
	Partially Meeting Learning Standards	135.3	0.9	136	147	18
	Meeting Learning Standards	147.7	1.0	148	188	69
	Meeting Learning Standards with Distinction	188.1	2.6	189	225	4
6	Not Meeting Learning Standards	NA	NA	100	130	5
	Partially Meeting Learning Standards	130.4	1.2	131	146	14
	Meeting Learning Standards	146.6	1.5	147	186	73
	Meeting Learning Standards with Distinction	186.1	0.9	187	225	8
7	Not Meeting Learning Standards	NA	NA	100	132	10
	Partially Meeting Learning Standards	132.6	1.6	133	160	62
	Meeting Learning Standards	160.4	2.9	161	195	25
	Meeting Learning Standards with Distinction	195.7	2.2	196	225	3
8	Not Meeting Learning Standards	NA	NA	100	135	12
	Partially Meeting Learning Standards	135.2	1.6	136	167	73
	Meeting Learning Standards	167.6	4.1	168	193	12
	Meeting Learning Standards with Distinction	193.5	1.3	194	225	3
HS	Not Meeting Learning Standards	NA	NA	100	131	6
	Partially Meeting Learning Standards	131.3	0.9	132	151	28
	Meeting Learning Standards	151.6	3.0	152	194	62
	Meeting Learning Standards with Distinction	194.2	3.2	195	225	5

Table 3-10. 2014 NYSAA Standard Setting: Round 3 Results—Mathematics

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
3	Not Meeting Learning Standards	NA	NA	100	137	9
	Partially Meeting Learning Standards	137.5	1.6	138	170	57
	Meeting Learning Standards	170.2	0.8	171	200	30
	Meeting Learning Standards with Distinction	200.5	1.0	201	225	3
4	Not Meeting Learning Standards	NA	NA	100	135	12
	Partially Meeting Learning Standards	135.2	1.1	136	161	48
	Meeting Learning Standards	161.4	0.6	162	192	34
	Meeting Learning Standards with Distinction	192.9	1.9	193	225	7
5	Not Meeting Learning Standards	NA	NA	100	127	7
	Partially Meeting Learning Standards	127.5	0.0	128	143	18
	Meeting Learning Standards	143.5	0.0	144	192	72
	Meeting Learning Standards with Distinction	192.1	0.6	193	225	3
6	Not Meeting Learning Standards	NA	NA	100	128	6
	Partially Meeting Learning Standards	129.0	0.0	129	147	24
	Meeting Learning Standards	148.0	0.0	148	183	59
	Meeting Learning Standards with Distinction	184.0	0.0	184	225	11
7	Not Meeting Learning Standards	NA	NA	100	124	3
	Partially Meeting Learning Standards	124.6	0.4	125	158	49
	Meeting Learning Standards	158.1	0.7	159	194	41
	Meeting Learning Standards with Distinction	194.6	0.7	195	225	7
8	Not Meeting Learning Standards	NA	NA	100	132	10
	Partially Meeting Learning Standards	132.4	0.6	133	159	59
	Meeting Learning Standards	159.4	2.2	160	206	30
	Meeting Learning Standards with Distinction	206.2	2.3	207	225	2
HS	Not Meeting Learning Standards	NA	NA	100	129	6
	Partially Meeting Learning Standards	129.1	1.0	130	141	9
	Meeting Learning Standards	141.2	1.2	142	201	82
	Meeting Learning Standards with Distinction	201.4	4.4	202	225	3

Table 3-11. 2014 NYSAA Standard Setting: Round 3 Results—Science

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
4	Not Meeting Learning Standards	NA	NA	100	109	3
	Partially Meeting Learning Standards	109.4	0.3	110	117	18
	Meeting Learning Standards	117.7	1.0	118	142	78
	Meeting Learning Standards with Distinction	142.6	2.2	143	150	1
8	Not Meeting Learning Standards	NA	NA	100	112	6
	Partially Meeting Learning Standards	112.3	0.2	113	121	30
	Meeting Learning Standards	121.1	1.1	122	143	61
	Meeting Learning Standards with Distinction	143.8	0.9	144	150	3
HS	Not Meeting Learning Standards	NA	NA	100	109	3
	Partially Meeting Learning Standards	110.0	0.8	110	121	34
	Meeting Learning Standards	121.6	0.6	122	139	51
	Meeting Learning Standards with Distinction	139.2	0.4	140	150	12

Table 3-12. 2014 NYSAA Standard Setting: Round 3 Results—Social Studies

Grade	Performance Levels	Average Cut	Standard Error	Score Range		Percent of Students
				Minimum	Maximum	
HS	Not Meeting Learning Standards	NA	NA	100	108	3
	Partially Meeting Learning Standards	109.0	1.0	109	120	25
	Meeting Learning Standards	120.9	1.5	121	136	64
	Meeting Learning Standards with Distinction	136.7	1.6	137	150	7

3.12 SEQUENCE OF GRADE LEVELS

The process described in the previous paragraphs was followed for each grade/content area. Each panel was responsible for recommending standards for two grade levels, with the exception of those recommending standards at the high school level, those panelists only recommended standards for a single grade and content area. At the high school level, those panelists set standards for only a single grade and content area. See Table 1-1 for the configuration of the panels. Therefore, the results presented in Tables 3.8 through 3.11 represent a repetition of the process by each panel. In each case, a panel would complete the process for their first grade level, starting with the review of the assessment materials and ending with the Round 3 ratings, and then repeat the entire process one more time for the remaining grade level.

3.13 EVALUATION

The measurement literature sometimes considers the evaluation process to be another product of the standard-setting process (e.g., Reckase, 2001), as it provides important validity evidence supporting the cut scores that are obtained. To provide evidence of the participants' views of the standard-setting process, panelists were asked to complete the evaluation about the general session presentations, the training round, and also about the standard-setting process itself. The results of the evaluations are presented in Appendix I.

CHAPTER 4 TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING

Upon conclusion of the standard-setting meeting, several important tasks were completed. These tasks centered on the following: convening a cross-grade articulation committee to review the cut scores for all grades and content areas except for social studies, as it is only one grade level; reviewing the standard-setting process and addressing issues presented by the outcomes; presenting the results to the New York State Education Department (the Department); making any final revisions or adjustments based on policy considerations under the direction of the Department; and preparing the standard-setting report.

4.1 CROSS-GRADE ARTICULATION COMMITTEE MEETING

Upon completion of the standard-setting process, a cross-grade articulation committee was convened. Two panelists from each of the English Language Arts (ELA), mathematics, and science panels were asked to be a part of this meeting. Panelists were given an overview of their tasks, which were to review the cuts and impact data for each of the content areas, discuss with the group any concerns or observations that they had about the data, and fill out a cross-grade evaluation (a sample of the evaluation is included in Appendix J). With the implementation of the new test design of the New York State Alternate Assessment (NYSAA), the Department determined that there would be guardrails placed around the third cut score such that the percent of students in the top two performance categories (Meeting Learning Standards and Meeting Learning Standards with Distinction) would be at least 70%. The articulation panel was instructed to adjust the cut scores so that this minimum percentage was met. Once the minimum percentage was met, they discussed the remaining two cut scores and decided to have the percentage of students in all four performance categories similar across the grades. The adjustments were greatest for mathematics Grades 3/4, where the cuts resulting from the standard setting led to far fewer than 70% of students Meeting Learning Standards. Discussions between the panelists revealed that the groups setting the standards in these grades used much more stringent criteria than those in other grades/contents. Adjustments were made in the cut scores to be more in line with the other grade/content areas.

4.2 ANALYSIS AND REVIEW OF PANELISTS' FEEDBACK

Upon completion of the evaluation forms, panelists' responses were reviewed by the psychometrician. This review did not reveal any anomalies in the standard-setting process. In general, participants were happy but their cuts were changed later.

4.3 POLICY ADJUSTMENTS

After all of the standard-setting activities had been completed and all of the materials were reviewed, the final cut scores recommended by the panelists in the articulation panel were accepted by the Commissioner. Final cut scores can be found in Appendix K.

4.4 PREPARATION OF STANDARD SETTING REPORT

Following the final compilation of the standard-setting results, Measured Progress prepared this report, which documents the procedures and results of the 2014 standard-setting meeting in order to establish performance standards for the NYSAA in ELA, mathematics, science, and social studies.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- Ferrara, S., Swaffield, S., & Mueller, L. (2009). Conceptualizing and setting performance standards for alternate assessments. In W. D. Schafer & R. W. Lissitz (Eds.), *Alternate Assessments Based on Alternate Achievement Standards: Policy, Practice and Potential* (pp. 93–111). Baltimore, MD: Paul H. Brookes Publishing Co.
- Kingston, N. K., Kahl, S. K., Sweeney, K. P., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 219–248). Mahwah, NJ: Lawrence Erlbaum.
- Measured Progress. (2010). *2009–10 New Hampshire alternate assessment technical report*. Unpublished report.
- Reckase, M. D. (2001). Innovative methods for helping standard setting participants to perform their task: The role of feedback regarding consistency, accuracy, and impact. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 159–173). Mahwah, NJ: Lawrence Erlbaum.