

**NEW YORK STATE ALTERNATE ASSESSMENT
(NYSAA)**

STANDARD SETTING REPORT

**ENGLISH LANGUAGE ARTS AND MATHEMATICS,
GRADES 3–8 AND HIGH SCHOOL
SCIENCE, GRADES 4, 8, AND HIGH SCHOOL
SOCIAL STUDIES, GRADES 5, 8, AND HIGH SCHOOL**

JUNE 9–13, 2008



Developed by
Measured Progress, Inc.

Table of Contents

INTRODUCTION	1
OVERVIEW OF STANDARD SETTING	2
OVERVIEW OF PROCESS	3
TASKS COMPLETED PRIOR TO STANDARD SETTING	4
CREATION OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS.....	4
SELECTION OF STUDENT DATAFOLIOS.....	4
PREPARATION OF MATERIALS FOR PANELISTS	6
PREPARATION OF PANELIST TRAINING PRESENTATIONS	6
PREPARATION OF INSTRUCTIONS FOR FACILITATORS DOCUMENT	6
PREPARATION OF SYSTEMS AND MATERIALS FOR ANALYSIS DURING THE SESSION.....	6
SELECTION OF PANELISTS.....	7
TASKS COMPLETED DURING STANDARD SETTING	9
ORIENTATION.....	9
REVIEW OF ALTERNATE PERFORMANCE LEVEL DESCRIPTORS	9
ROUND 1 JUDGMENTS FOR MIDDLE CUT	10
ROUND 2 JUDGMENTS FOR MIDDLE CUT	10
ROUND 1 AND 2 JUDGMENTS FOR LOWER CUT	10
ROUND 1 AND 2 JUDGMENTS FOR UPPER CUT.....	10
TABULATION OF ROUND 2 RESULTS.....	11
ROUND 3 JUDGMENTS	11
RECOMMENDATIONS FOR CHANGES TO ALTERNATE PERFORMANCE LEVEL DESCRIPTORS	14
REPETITION OF STANDARD SETTING PROCESS FOR SECOND GRADE LEVEL	14
EVALUATION OF STANDARD SETTING PROCESS BY PANELISTS.....	14
POLICY FORUM	14
TASKS COMPLETED AFTER STANDARD SETTING	18
ANALYSIS AND REVIEW OF PANELISTS’ FEEDBACK	18
CALCULATION OF RECOMMENDED CUT SCORES	18
PREPARATION OF THIS STANDARD SETTING REPORT	21
REFERENCES	22
APPENDICES	24
<i>Appendix A—Meeting Agenda</i>	25
<i>Appendix B—Alternate Performance Level Descriptors</i>	27
<i>Appendix C—Opening Session PowerPoint Presentations</i>	48
<i>Appendix D—Group Facilitator’s Scripts and Outlines</i>	72
<i>Appendix E—List of Participants</i>	99
<i>Appendix F—Sample Rating Forms</i>	103
<i>Appendix G—Sample Evaluation Form</i>	108
<i>Appendix H—Results of Evaluation by Panelists</i>	111
<i>Appendix I—Policy Forum Agenda</i>	147
<i>Appendix J—Policy Forum PowerPoint Presentation</i>	149

Appendix K—Policy Forum Tables and Graphs 153
Appendix L—Policy Forum Tables and Graphs of Recommendations 166
*Appendix M—Sample Evaluation Form and Results of Policy Forum Evaluation by
Panelists* 179

INTRODUCTION

In response to requirements outlined in the Individuals with Disabilities Education Act (IDEA) Amendments of 1997, the reauthorization of IDEA in 2004, and the No Child Left Behind Act of 2001 (NCLB), states have developed alternate assessments for students with disabilities. A variety of measurement formats have been implemented in these assessment systems (Thompson & Thurlow, 2001; Roeber, 2002; Smith, 2003; Malehorn, 1994; Navarrete, Wilde, Nelson, Martinez, & Hargett, 1990). Due to differential requirements within their Individual Education Programs (IEPs), students with disabilities may be administered different assessments appropriate to their level of ability. The test scores and performance level categories of these students, however, are reported as a single group. Given the nature of the alternate assessments, setting performance level standards for the alternate assessments can be challenging in terms of educational and policy considerations.

A number of standard setting methods have been developed over the last 30 years (Berk, 1986; Reckase, 2000; Hambleton, Jaeger, Plake, & Mills, 2000; Cizek, 2001; Hambleton & Powell, 1983; Kane, 1994; Livingston & Zieky, 1982; Lunz, 1995). Most of the methods (e.g., Bookmark, Body of Work, etc.) were developed in large-scale assessment settings. Each has its advantages as well as a number of limitations. The choice for a particular application should be based on a thorough review of existing methods in terms of their pros and cons for the concrete testing situation at hand (Cizek, 1996; Reckase, 2000; Hambleton, 2001). The most important criteria are:

- (a) The appropriateness of the method for the concrete situation.
- (b) The feasibility of the method implementation under the current circumstances.
- (c) The existing validity evidence for the quality of the selected method.

Given the complexity of alternate assessments (e.g., differential assessments, unique learning attributes of this population, etc.), there is increased emphasis on developing new standard setting methods or modifying existing methods appropriate to these new conditions. Not many methods can address the complexity, so states tend to retrofit existing methods to their alternate assessment programs. Some of the very popular standard setting methods used in alternate assessment programs so far include Modified Angoff (Angoff, 1971), Bookmark (Lewis, Mitzel, & Green, 1996), Body of Work (Kingston, Kahl, Sweeney, & Bay, 2001), and Judgmental Policy Capturing (Jaeger, 1995).

Feasibility and validity are of great importance when evaluating a standard setting method (Cizek, 1996). The modified Body of Work (mBoW) procedure was chosen for the New York State Alternate Assessment (NYSAA). This method is based on the Body of Work (BoW) method described in Kingston, Kahl, Sweeney, & Bay (2001). The methodological modification for BoW is one that is appropriate for a portfolio type of assessment such as NYSAA. In this method, panelists review student datafolios that represent the range of student scores. The panelists independently classify each student datafolio into one of four following performance levels based on their understanding of the alternate performance level descriptors:

- *Meets the Learning Standards with Distinction*
- *Meets the Learning Standards*
- *Partially Meets the Learning Standards*
- *Does Not Meet the Learning Standards*

Overview of Standard Setting

During the summer of 2006, the New York State Education Department, in collaboration with Measured Progress, redesigned New York State Alternate Assessment (NYSAA). The purpose of the redesign was to ensure that students with severe cognitive disabilities were being assessed against grade level content. The content was expanded to Alternate Grade Level Indicators (AGLIs) which are aligned to the general education grade level indicators. The redesigned NYSAAs were administered during the 2006–07 school year. Because the assessment was new to students, teachers, and administrators, the Department and Measured Progress held a stakeholders meeting to refine the process even further. The 2007–08 administration was the conclusion of a two-year endeavor to redesign, implement, and set standards; then refine, implement, and conduct a final standard setting in early summer of 2008. With this major shift in NYSAAs administration, the Department conducted the first of two standard setting sessions during June 2007 for the 2006–07 administration. The 2007 standard setting was held to establish cut scores for English language arts and mathematics in grades 3 through 8 and high school, science in grades 4, 8, and high school, and social studies in grades 5, 8, and high school.

The 2007 standard setting method implemented for all grade levels and content areas was the modified Performance Profile procedure. This procedure entailed two separate reviews of all possible student score combinations. During the first stage, panelists reviewed 256 possible score combinations for a single data collection period. A data collection period for the purposes of the standard setting process was defined as a single date of recorded student performance data documented in a datafolio. The panelists had the opportunity to rate each combination based on accuracy and independence. The second round of ratings was based on the required three data collection periods. To assist panelists with understanding datafolio contents, they had the opportunity to review a sample of student datafolios. This allowed panelists to understand the student knowledge, skills, and understandings that were represented across all four performance levels. Panelists also were required to discuss and come to consensus on the Alternate Performance Level Descriptors (APLDs). A complete description of the 2007 standard setting procedures can be found in the standard setting report produced for the June 2007 modified Performance Profile procedure.

For the 2007–08 administration, the Department was provided with a standard setting proposal that outlined the mBoW as plan A and the enhanced Performance Profile procedure (eP³) as plan B. Both the Department and the New York State Technical Advisory Group preferred to implement plan A for the 2008 standard setting. In May 2008, it was determined that plan A could be implemented as Measured Progress received the full set of 20 percent of NYSAAs datafolios from each scoring institute.

Standard setting to establish cut scores for NYSAAs was conducted Monday through Friday, June 9–13, 2008. The meetings took place at the Crowne Plaza Hotel in Albany, New York. For English language arts and mathematics, six of the eight standard setting panels established cut scores for two grade levels (3 and 4, 5 and 6, or 7 and 8); for science and social studies, two of the four panels established cuts for two grade levels (8 and high school); and the remaining four panels (English language arts and mathematics high school, science grade 4, and social studies grade 5) recommended cut scores for a single grade level.

The mBoW method that was used in 2008 is basically a sorting task, in which panelists sort a set of student datafolios (bodies of work) into four piles corresponding to the performance levels. First, the panelists divided the datafolios into two piles, corresponding to datafolios representing performance below *Meets the Learning Standards* and those representing *Meets the Learning Standards* or above, and then each of those piles was further subdivided into two piles. A step-by-step description of the method is provided in the next section of this report. All panels followed the same procedures.

To help ensure consistency of procedures among panels, all panelists received initial training as a group, and then each panel was led through the standard setting process by a trained facilitator from Measured Progress.

Overview of Process

This section provides an overview of the standard setting process as it was implemented for NYSAA. The process was divided into three stages, each with several constituent tasks.

❖ Tasks completed prior to standard setting:

- Creation of Alternate Performance Level Descriptors (APLDs)
- Selection of student datafolios
- Preparation of materials for panelists
- Preparation of panelist training presentation
- Preparation of Instructions for Facilitators document
- Preparation of systems and materials for analysis during the sessions
- Selection of panelists

❖ Tasks completed during standard setting:

- Orientation
- Review of APLDs
- Sorting of datafolios into performance level categories
- Round 1 judgments for middle cut
- Round 2 judgments for middle cut
- Round 1 and 2 judgments for lower cut
- Round 1 and 2 judgments for upper cut
- Tabulation of Round 2 results
- Round 3 judgments for all three cuts
- Recommendations for changes to APLDs
- Repetition of standard setting process for second grade level (English language arts and mathematics, grades 4, 6, and 8; science and social studies, high school)
- Evaluation of standard setting process by panelists
- Policy forum

❖ Tasks completed after standard setting:

- Analysis and review of panelists' feedback
- Calculation of recommended cut scores
- Preparation of this standard setting report

Tasks Completed Prior to Standard Setting

Creation of Alternate Performance Level Descriptors

The Alternate Performance Level Descriptors (APLDs) provided 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 are included in Appendix B.

Selection of Student Datafolios

The student datafolios were selected from the sample received from across the state for Reliability Monitoring Review (RMR). Measured Progress receives approximately 20 percent of NYSAA datafolios from across the state. This 20 percent sample is representative of all students who take the alternate assessment. The 20 percent RMR sample represented the grades and content areas assessed for NYSAA. The table below outlines the requirements for NYSAA

Content Requirements for Each Grade

Content Area	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	High School
English Language Arts	X	X	X	X	X	X	X
Mathematics	X	X	X	X	X	X	X
Science		X				X	X
Social Studies			X			X	X

Each datafolio administered as part of NYSAA includes two required components. Within each required component are two choice components. This is based on the NYSAA test blueprint. The following is an example of the required and choice components from the test blueprint for English language arts (ELA).

REQUIRED COMPONENT

Two ELA Key Ideas Must be Assessed at each Grade Level

Required Key Ideas Vary by Grade as Marked by an X in the Chart Below

ELA Key Idea	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 COMPONENT

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 in the Chart Below Choose 1 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 I
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						

Within each choice component, the student is assessed against one Alternate Grade Level Indicator (AGLI). Data on accuracy and independence is collected on three different dates during the administration period of the assessment. Accuracy and independence have a score point ranging from 1 to 4 for a single date. For example, in English language arts grade 3, the student receives a score for reading accuracy, reading independence, listening accuracy, and listening independence.

Table 1:
Example of Score Range for a Single Grade and Content Area

	Date 1	Date 2	Date 3	Total
Required Component 1				
Level of Accuracy	1-4	1-4	1-4	3-12
Level of Independence	1-4	1-4	1-4	3-12
Required Component 2				
Level of Accuracy	1-4	1-4	1-4	3-12
Level of Independence	1-4	1-4	1-4	3-12

The total score from the three dates of student performance for the ratings of level of accuracy and level of independence were computed based on the decision rules that were provided by the Department. Because the thresholds were based on the panelists' classification of student performance, the selection of datafolios to be used in the process

was a crucial part of the preparation. One set of 20 to 25 student datafolios representing a range of student performance scores from 12 to 48 was the desired goal. The student data file of responses was sorted by overall grade/content scores. A range of response intervals was determined (e.g., every 4 to 5 score points), and a randomizing program was used to select student responses within each of the intervals. These datafolios were then pulled, scanned, prepped, and copied by Measured Progress for the standard setting process. If a datafolio could not be copied or was otherwise unsuitable for use (i.e., incomplete student performance scores), a replacement response, if possible, was selected. Because the score distributions are skewed, with far larger numbers of students receiving scores at the high end of the score scale, there were not always many datafolios representing the lower scores available for use. The numbers of datafolios selected ranged from 20 to 23, across the grades and content areas.

Preparation of Materials for Panelists

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

- Meeting agenda
- Nondisclosure agreement
- APLDs
- Set of 20 to 23 datafolios
- Set of four rating forms
- Evaluation form

Copies of the meeting agenda, APLDs, sample rating forms, and the evaluation form are included in the appendices.

Preparation of Panelist Training Presentations

The PowerPoint presentations used in the opening session was prepared prior to the meeting and was designed to give panelists a preview of what to expect throughout the standard setting process. The PowerPoint presentations are included in Appendix C.

Preparation of Instructions for Facilitators Document

A document, “General Instructions for NYSAA Standard Setting Group Facilitators,” was created for the group facilitators to refer to as they worked through the process, and to help ensure uniformity during implementation of the various tasks across the grades and content areas. Separate versions were created for groups setting standards for a single grade level and those setting standards for two grade levels. In addition, outline versions of the facilitator’s instructions were created. Copies of the facilitator’s scripts and outlines are included in Appendix D.

Preparation of Systems and Materials for Analysis During the Session

Programming of all analyses to be conducted during the standard setting sessions was completed and thoroughly tested prior to the start of the session. See additional information about the statistical analyses in the “Tabulation of Round 2 Results” section on page 10 of this report.

Selection of Panelists

Panelists were recruited and selected prior to the standard setting session by the Department. The goal was to recruit approximately 10 people for each panel, with a total of 120 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. It was recommended that each panel be 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, it was recommended that the selection of panelists should reflect 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 the grade(s) for which they would be setting standards.

Panelists were recruited through various groups and organizations. The list below includes several of the groups and organizations that were contacted for panel participation:

- New York State Alternate Assessment Advisory Committee
- New York State Alternate Assessment Training Network (AATNs)
- New York State Alternate Assessment Score Site Coordinators (SSCs)
- New York State Alternate Assessment Revision Work Group (NRWG)
- New York State Alternate Assessment Regional Lead Trainers (RLTs)
- 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 (CSE) throughout New York State
- Board of Cooperative Educational Services (BOCES)
- New York State Operated and State Approved Schools
- New York State United Teachers (NYSUT)
- New York State Education Department
 - Office of State Assessment
 - Office of Curriculum Instruction and Instructional Technology
 - Office of Vocational and Educational Services for Individuals with Disabilities
- 2006-07 New York State Alternate Assessment Standard Setting Panelists
- General and Special Education Teachers throughout New York State

The actual number of panelists who participated was 95, distributed into 12 panels as shown in Table 2.

Table 2:
Number of Panelists per Group

Grade Level	English Language Arts	Mathematics	Science	Social Studies
3	9	8	9	
4				
5	7	9		8
6				
7	9	7		
8				
High School	7	7	7	8

Eleven panelists identified themselves as being members of the Alternate Assessment Training Network (AATN), six indicated they were Score Site Coordinators (SSCs), and an additional nine were in both categories. All panelists had at least a bachelor's degree, and 93 of the 95 had an advanced degree as well (87 had master's degrees and 6 had doctoral degrees). Sixty-three panelists indicated they had experience with students with severe cognitive disabilities, and 62 had administration or training experience with NYSAA. The panelists' years of teaching experience ranged from 2 to 40 and averaged about 20 years.

Tables 3 through 8 show some additional information about the demographic background and work experience of the panelists. Note that, for Tables 3 through 6 and 8, some panelists selected more than one of the options, so the numbers sum to more than 95.

Table 3:
Panelists by Job Title

Title	Number
Special Education Teacher	39
General Education Teacher	21
Administrator	19
Other*	22
*(Some responses in the Other category include: Vocational Evaluator, Autism Specialist, Staff Developer, Reading Specialist, Retired Content Teacher, Retired Special Education Teacher, Testing Consultant—DAA, AIS English/Reading Teacher, AIS Remedial Mathematics Teacher, etc.)	

Table 4:
Panelists by Level of School Experience

Level	Number
Elementary	57
Intermediate	49
High School	33

Table 5:
Panelists by Content Area School Experience

Content Area	Number
English Language Arts	34
Mathematics	29
Science	16
Social Studies	17

Table 6:
Panelists by Community Type

Community Type	Number
Rural	17
Urban	34
Suburban	49

Table 7:
Panelists by Gender

Gender	Number
Female	76
Male	19

Table 8:
Panelists by Ethnicity

Ethnicity	Number
Asian	2
Black/African American	4
Hispanic/Latino	3
White	86

A list of the standard setting panelists is included in Appendix E.

Tasks Completed During Standard Setting

Orientation

The standard setting session began with a general orientation that was attended by all panelists. The purpose of the orientation was to provide some background information and an introduction to the issues of standard setting, and to explain the activities that would occur during standard setting. Panelists were also given an overview of NYSAA and how it is scored. At the conclusion of the opening session, panelists were given an opportunity to ask questions about the standard setting process. After the large group session, the panelists assembled into their content area/grade level groups in separate rooms.

Review of Alternate Performance Level Descriptors

The first activity once the panelists were in their breakout rooms was to review the Alternate Performance Level Descriptors (APLDs). This important step of the process was designed to ensure that the panelists had a thorough and common understanding of the knowledge, skills, and understandings students needed to demonstrate in order to be classified as *Partially Meets the Learning Standards*, *Meets the Learning Standards*, and *Meets the Learning Standards with Distinction*. The panelists began by individually reviewing the APLDs, and then they discussed them as a group, coming to consensus as to the characteristics that define a student at each level. As a group, the panelists created bulleted lists of characteristics that describe students at each performance level; the lists

were posted in the room on chart paper for the panelists to refer to as they worked through the three rounds of ratings. The APLDs are provided in Appendix B.

Round 1 Judgments for Middle Cut

For Round 1, panelists worked individually, without discussion with their colleagues. The panelists began by individually reviewing the full set of datafolios. As the panelists worked through the pile of datafolios, they kept in mind the APLDs, in particular, the set of knowledge, skills, and understandings students need to demonstrate in order to be classified as *Meets the Learning Standards* or above. Once they completed the review of the full set of datafolios, panelists started again with the first datafolio and separated the datafolios into two piles: those representing performance below *Meets the Learning Standards* versus those representing performance at *Meets the Learning Standards* or above. Once they had separated the datafolios into two piles, the panelists completed the “Round 1–Individual” section of the middle cut rating form.

Round 2 Judgments for Middle Cut

After all panelists had completed their Round 1 ratings, the group facilitator recorded the Round 1 ratings on chart paper using a show of hands. The chart showed the number of panelists who assigned each datafolio to each of the two piles. Next, the panelists discussed their Round 1 ratings, starting with the first datafolio for which there was disagreement as to how it should be categorized. They progressed through the full set of datafolios, discussing their rationale for how they categorized each. The facilitator told them they did not need to come to consensus as to how the datafolios were categorized, but that they should both express their own viewpoints and listen to those of their colleagues. Once the discussions were complete, the panelists filled out the “Round 2–After Discussion” section of the middle cut rating form.

Round 1 and 2 Judgments for Lower Cut

Once panelists completed Round 2 for the middle cut and had separated the datafolios into two piles, they then repeated the process for the lower cut. The panelists set the pile of datafolios categorized as *Meets the Learning Standards* or above aside, then, working individually and beginning with the first datafolio in the below *Meets the Learning Standards* pile, reviewed each datafolio in turn. As they worked through the pile, the panelists placed each datafolio into one of two new piles, corresponding to *Does Not Meet the Learning Standards* versus *Partially Meets the Learning Standards*. When they had completed their review of all the below *Meets the Learning Standards* datafolios, the panelists completed the Round 1 section of the lower cut rating form.

As with the middle cut, the facilitator then summarized the Round 1 ratings on chart paper using a show of hands, and the panelists discussed their categorizations, starting with the first datafolio for which there was disagreement as to how it should be categorized. The facilitator reminded the panelists that they did not need to come to consensus. Once the discussions were complete, the panelists filled out the Round 2 section of the lower cut rating form.

Round 1 and 2 Judgments for Upper Cut

Finally, the panelists repeated Rounds 1 and 2 one last time for the upper cut. The datafolios classified as *Does Not Meet the Learning Standards* and *Partially Meets the*

Learning Standards were set aside, and the panelists began going through each of the datafolios in the above *Meets the Learning Standards* pile. As they worked through the pile, the panelists placed each datafolio into one of two new piles, corresponding to *Meets the Learning Standards* versus *Meets the Learning Standards with Distinction*. As with the previous two cuts, the panelists completed Round 1 individually, without discussion with their colleagues, filled out the Round 1 section of the upper cut rating form, and then discussed their ratings as a group and completed the Round 2 section of the form.

Tabulation of Round 2 Results

After the panelists completed their Round 2 ratings for all three cuts, the rating forms were submitted to the data analysis work room, where the data entry and analysis of the round 2 ratings were done. The following information was calculated and shared with the panelists:

1. The Round 2 average cut scores, based on the ratings of all panelists. Cuts were calculated using SAS statistical software Version 9.1 (SAS Institute Inc., 2004) by first determining each panelist's individual raw score cuts using logistic regression (PROC LOGISTIC), then taking the mean across panelists to get the overall average cuts (PROC MEANS).
2. Impact data, showing the percentage of students across the state who would fall into each performance level according to the Round 2 average cuts. The impact data were based on the total population of students who took the NYSAA in 2007-08 administration.
3. The individual panelists' cuts for each cut score. The individual cuts on the raw score scale were shown on a chart using an alphabetic code for each panelist. The purpose of providing the individual panelist cuts was to allow the panelists to see to what extent their ratings agreed or disagreed with the rest of the panelists' ratings. Even though the panelists did not need to come to consensus, this step enabled panelists who were outliers to self-correct at their option.

Round 3 Judgments

A psychometrician presented the Round 2 feedback to the panelists and explained the process for the Round 3 ratings. The Round 3 discussions were conducted similarly to earlier discussions, with the facilitator summarizing the Round 2 ratings on chart paper and the panelists discussing any datafolios for which there was disagreement. In this case, however, the panelists were encouraged to look at all three cuts together. In addition, they were told to consider the impact data as a reasonableness check on the placement of the cuts. The psychometrician was careful to tell the panelists that they shouldn't consider the impact data in isolation, but instead use it only to identify any potentially problematic cuts. The psychometrician reiterated that changes in the ratings should always be based on the APLDs and the knowledge, skills, and understandings demonstrated by the students in their datafolios. Once the discussions were completed for the full set of datafolios, the panelists entered their final ratings on the Round 3 rating form. The Round 3 cuts, calculated as described above for Round 2, represented the standard setting panelists' final recommendations.

Tables 9 through 12 show the average cuts and standard errors for Rounds 1, 2 and 3. Examination of the columns headed "Mean Cut" shows the extent to which the panelists' ratings changed from one round to the next as a result of the discussions. In general,

changes in the average cuts were fairly small across rounds, although there were some exceptions. Also shown in Tables 9 through 12 are standard errors of measurement. These values were obtained by calculating the standard deviation of the panelists' ratings and dividing by the square root of the number of panelists ($\frac{\sigma_p}{\sqrt{n_p}}$). Examination of the standard

errors in Tables 9 through 12 shows that, in general, the values tended to get smaller with each round of ratings, indicating increased agreement among the panelists as a result of the rounds of discussion. Note that the standard errors were not used as part of the standard setting process; they are provided here for informational purposes.

Table 9: English Language Arts – Average Cuts and Standard Errors of Measurement by Round

Grade	Cut	Round 1		Round 2		Round 3	
		Mean Cut	Std. Error	Mean Cut	Std. Error	Mean Cut	Std. Error
3	1	29.6	1.99	30.9	0.66	32.1	1.03
	2	38.5	1.59	38.6	0.90	39.2	0.91
	3	45.8	0.25	45.5	0.25	44.2	1.46
4	1	34.5	0.60	31.2	0.93	32.7	0.81
	2	41.7	0.48	40.7	0.29	40.5	0.32
	3	45.8	0.31	45.5	0.00	45.8	0.28
5	1	29.5	0.00	29.5	0.00	29.5	0.00
	2	37.8	0.46	37.8	0.46	38.5	0.00
	3	44.3	1.10	46.6	0.21	46.9	0.14
6	1	27.4	1.51	27.5	0.00	27.5	0.00
	2	38.4	0.86	38.4	0.86	40.4	0.65
	3	46.7	0.21	46.5	0.00	45.9	0.57
7	1	23.7	1.80	22.5	1.28	21.8	1.18
	2	36.3	0.67	36.3	0.51	36.4	0.36
	3	41.7	0.91	41.4	0.56	42.0	0.33
8	1	27.3	1.58	24.0	1.50	22.0	1.00
	2	34.8	0.10	34.7	0.08	34.0	0.38
	3	40.9	0.36	40.9	0.53	42.1	0.49
High School	1	29.5	4.00	24.9	2.21	28.4	2.42
	2	34.3	0.76	34.5	0.46	34.5	0.46
	3	42.7	1.14	42.0	0.50	42.0	0.50

Table 10: Mathematics – Average Cuts and Standard Errors of Measurement by Round

Grade	Cut	Round 1		Round 2		Round 3	
		Mean Cut	Std. Error	Mean Cut	Std. Error	Mean Cut	Std. Error
3	1	26.5	1.54	24.3	1.19	22.5	0.00
	2	35.6	0.42	35.6	0.42	35.0	0.05
	3	43.7	1.05	45.3	0.06	45.6	0.19
4	1	29.5	1.91	23.0	0.00	23.0	0.00
	2	38.4	0.64	38.4	0.64	38.4	0.64
	3	45.9	0.49	46.2	0.17	46.5	0.25
5	1	28.9	1.12	26.4	1.46	24.5	1.40
	2	36.0	0.00	36.0	0.00	36.0	0.00
	3	44.5	0.63	46.4	0.88	46.4	0.88
6	1	23.3	1.04	21.4	0.64	21.4	0.64
	2	33.7	0.31	33.4	0.37	33.4	0.37
	3	43.4	0.41	43.4	0.37	43.2	0.42
7	1	32.7	0.53	33.1	0.36	33.4	0.14
	2	36.8	0.53	36.7	0.46	37.1	0.51
	3	46.0	0.58	46.2	0.41	46.9	0.24
8	1	30.4	1.01	32.0	0.00	32.0	0.00
	2	38.4	0.48	38.4	0.48	38.4	0.48
	3	46.6	0.44	47.5	0.00	47.5	0.00
High School	1	22.0	3.70	20.4	0.35	20.0	0.00
	2	33.8	0.71	34.0	0.60	34.7	0.90
	3	44.9	0.56	45.1	0.40	45.1	0.40

Table 11: Science – Average Cuts and Standard Errors of Measurement by Round

Grade	Cut	Round 1		Round 2		Round 3	
		Mean Cut	Std. Error	Mean Cut	Std. Error	Mean Cut	Std. Error
4	1	29.5	1.64	29.7	1.46	29.8	1.48
	2	33.9	0.85	34.0	0.76	34.0	0.76
	3	44.1	0.67	43.9	0.24	43.9	0.17
8	1	28.6	1.02	28.6	0.86	28.5	0.00
	2	38.1	0.44	38.1	0.44	38.5	0.00
	3	42.6	0.49	43.5	0.43	43.7	0.03
High School	1	30.3	0.08	21.0	0.00	21.0	0.00
	2	35.6	0.43	35.6	0.43	35.6	0.43
	3	45.2	0.71	43.5	0.00	43.1	0.36

Table 12: Social Studies – Average Cuts and Standard Errors of Measurement by Round

Grade	Cut	Round 1		Round 2		Round 3	
		Mean Cut	Std. Error	Mean Cut	Std. Error	Mean Cut	Std. Error
5	1	35.5	0.53	34.5	0.40	33.5	0.00
	2	40.2	0.00	39.6	0.32	41.0	0.00
	3	46.6	0.54	47.9	0.05	47.9	0.04
8	1	32.7	1.63	34.5	0.37	34.0	0.30
	2	36.2	0.11	36.1	0.10	36.4	0.14
	3	48.1	0.66	47.1	0.24	47.1	0.25
High School	1	32.2	0.31	32.5	0.00	32.5	0.00
	2	38.6	0.32	38.6	0.31	38.5	0.39
	3	47.1	0.51	46.5	0.31	46.3	0.35

Recommendations for Changes to Alternate Performance Level Descriptors

After completing the standard setting process, panelists in all groups were given an opportunity to provide suggested modifications or enhancements to the APLDs. Facilitators told the panelists that these were recommendations and that they might or might not be implemented.

Repetition of Standard Setting Process for Second Grade Level

After completing all of the steps for the first grade level, 8 of the 12 groups then repeated the entire process for the second grade level, starting with reviewing the APLDs, through all rounds of ratings, and ending with the recommendations for modifications to the APLDs.

Evaluation of Standard Setting Process by Panelists

As the last step in the standard setting process, panelists completed an evaluation form. The evaluation included a number of questions about the standard setting process, materials, etc., and gave the panelists an opportunity to indicate whether they believed the room average final cut scores were placed appropriately. These panelist judgments were analyzed to provide evidence of the procedural validity of the standard setting process. A sample evaluation form is included in Appendix G. Results of the evaluation are included in Appendix H.

Policy Forum

On Friday morning, after all the standard setting activities had been completed for all grades and content areas, a policy forum was held, attended by one representative from each grade/content area group. According to Cizek & Bunch (2007), “The goal of [the policy forum] is to yield a system of coherent and consistent cross-level and/or cross-content area performance standards” (p. 339). At the policy forum, participants were shown a set of tables and graphs—one for each content area—showing the results of the standard setting across the grade levels. The tables and graphs showed the raw score cuts as well as the impact data—the percentages of students scoring in a given performance level or higher—for each grade. Panelists were asked to discuss the overall results for each content area

and then make recommendations for any adjustments they felt should be made. The agenda for the policy forum is included in Appendix I; the PowerPoint presentation is in Appendix J; the tables and graphs showing the Round 3 standard setting results are in Appendix K; and the tables and graphs reflecting the policy forum participants' recommendations are in Appendix L. Steps 1 through 4 below outline the process used for the policy forum.

Step 1 - Welcome, Introductions, and Roles:

The group was welcomed and thanked for participating in the Policy Measurement Forum. Each attendee introduced themselves, indicated where they were from, gave some background information, and told which group they were representing. The roles and responsibilities of the Department and Measured Progress staff were reviewed, so each attendee understood who was doing what tasks during the forum. David Abrams, Assistant Commissioner for the Office of Standards, Assessment, and Reporting at the New York State Education Department, spoke to the attendees about some of the various reasons for holding the forum, provided some information about policy and measurement from the Department's perspective, and presented the overall charge for the meeting.

Step 2 - Review Standard Setting Purpose, How This Activity Supports Standard Setting, and All Materials Provided:

The forum facilitator reviewed the PowerPoint presentation and explained each of the sets of materials provided to the attendees. An opportunity for questions was provided. The facilitator asked the attendees to take some time to review the tables, charts, and graphs provided for further discussion of each. At this time it was also explained that while consensus was desired, it was not required for this process. Points of consensus as well as dissenting opinions were noted.

Step 3 - Review Each Content Area and Grade Cut Scores, and Seek Input:

Attendees were reminded that they were to look at each of the cut scores from a standard setting panelist viewpoint—taking into account the APLDs, their knowledge of the student population, and their knowledge of the assessment—as well as from a policy measurement viewpoint—taking into account the need for a fair and reasonable set of cut scores for NYSAA.

English Language Arts: Discussions began around each of the cut scores, particularly at the intermediate and high school grade levels. Each group representative provided a framework of reasons why the cut scores were set where they were based on discussions that occurred in each group during the standard setting process. The following adjustments were recommended:

- **High school:** *Partially Meets the Learning Standards* cut score was moved down from 29 to 27 because the general consensus was that it was more equitable given the other grade level cut scores at this performance level.
- Dissenting opinions were noted by a couple of attendees who felt that the cut score of 43 at grades 7, 8, and high school for *Meets the Learning Standards with Distinction* is too low given the APLDs and the nature of the assessment. Representatives explained that the

knowledge, skills, and understandings increase at these grade levels and felt that the cut score for this performance level should remain at 43.

Mathematics: Discussions began around each of the cut scores, particularly around the placement of cut scores at the intermediate grade levels, in which one *Meets the Learning Standards with Distinction* cut score was placed at 48 and where some of the ranges of raw scores within a performance level were too narrow. Each group representative provided a framework of reasons why the cut scores were set where they were based on discussions that occurred in each group during the standard setting process. The following adjustments were recommended:

- **Grade 6:** *Meets the Learning Standards* cut score was moved up from 43 to 44 and *Meets the Learning Standards with Distinction* was moved up from 44 to 45 because the full consensus was that it was more equitable given the APLDs and the other grade level cut scores at these performance levels. It also allowed for a better range of raw scores at the *Meets the Learning Standards* performance level.
- **Grade 7:** *Does Not Meet the Learning Standards* cut score was moved down from 33 to 31 and *Partially Meets the Learning Standards* cut score was moved down from 34 to 32 because full consensus was that it allowed for a better range of raw scores at the *Partially Meets the Learning Standards* performance level and provided a more consistent and fair cut score given various considered perspectives.
- **Grade 8:** *Meets the Learning Standards* cut score was moved down from 47 to 46 and *Meets the Learning Standards with Distinction* was moved down from 48 to 47 because the general consensus was that it was more equitable given the APLDs and the other grade level cut scores at these performance levels. Representatives expressed that the adjustment also equally took into account the policy and the group work perspectives.
- Dissenting opinions were noted by a couple of attendees who felt that the cut scores of 47 for the *Meets the Learning Standards with Distinction* performance level is still too high given the knowledge of the students who take NYSAA, the nature of the assessment, and the policy needs. Representatives explained that the APLDs and the reviewed datafolio samples at these grade levels were fair and representative of the students and felt that the cut scores for this performance level should remain at 47.

Science: Discussions began around each of the cut scores, particularly around the placement of cut scores at the grade 4 and high school grade levels for the performance levels of *Does Not Meet the Learning Standards* and *Partially Meets the Learning Standards*. Each group representative provided a framework of reasons why the cut scores were set where they were based on discussions that occurred in each group during the standard setting process. The following adjustments were recommended:

- **Grade 4:** *Does Not Meet the Learning Standards* cut score was moved down from 29 to 28 and *Partially Meets the Learning Standards* cut score was moved down from 30 to 29 because the full consensus was

that it was more equitable given the other grade level cut scores at these performance levels.

- **High school:** *Does Not Meet the Learning Standards* cut score was moved up from 20 to 23 and *Partially Meets the Learning Standards* cut score was moved up from 21 to 24 because the general consensus was that it was more equitable given the APLDs and the nature of the assessment. Representatives expressed that the adjustment also provided a more consistent range of raw scores given the other grade levels at these performance levels.
- No dissenting opinions were expressed by attendees for this content area.

Social Studies: Discussions began around each of the cut scores within the three grade levels for each of the four performance levels. Each group representative provided a framework of reasons why the cut scores were set where they were based on discussions that occurred in each group during the standard setting process. The following adjustments were recommended:

- **Grade 5:** *Meets the Learning Standards* cut score was moved down from 47 to 45 and *Meets the Learning Standards with Distinction* cut score was moved down from 48 to 46 because the general consensus was that it was more equitable given the nature of the students taking NYSAA and the nature of the assessment. Representatives expressed that the adjustment also equally took into account the policy and the group work perspectives.
- **Grade 8:** All performance level cut scores were adjusted to allow for a more equitable and consistent range of raw scores. *Does Not Meet the Learning Standards* was moved down from 33 to 31, *Partially Meets the Learning Standards* was moved down from 34 to 32, *Meets the Learning Standards* was moved down from 47 to 45, and *Meets the Learning Standards with Distinction* was moved down from 48 to 46. The general consensus was that each of the performance levels better met the APLDs, the student performance of the knowledge, skills, and understandings for NYSAA, and the nature of the assessment.
- **High school:** *Meets the Learning Standards* cut score was moved down from 46 to 45 and *Meets the Learning Standards with Distinction* was moved down from 47 to 46 because the general consensus was that it was more equitable given the other grade level cut scores at these performance levels.
- Dissenting opinions were noted by a couple of attendees regarding each of the other grades and content areas that had the cut score for *Meets the Learning Standards with Distinction* placed at 47. They felt that this cut score is still too high given the knowledge of the students who take NYSAA, the nature of the assessment, and the policy needs. They expressed that for each of the grades and content areas, the cut score of 47 should be adjusted to 46.

Each of the final recommended cut scores was recorded, and documentation for each of the content area cut scores was printed for the Department's records.

Step 4 - Wrap-up and Final Address to Attendees:

Forum attendees were asked to fill out the evaluation form. The evaluation and all materials were collected from each attendee by Measured Progress staff. The evaluation form and the results of the policy forum evaluation are presented in Appendix M. A final thank-you and explanation of the next step for the cut scores was provided by David Abrams, Assistant Commissioner.

Tasks Completed after Standard Setting

Upon conclusion of the standard setting sessions, several important tasks were completed. These tasks centered on reviewing the activities conducted during standard setting, addressing anomalies that may have occurred in the process or in the results, and compiling the recommended cut scores that resulted from the Round 3 ratings and the recommendations of the policy panel.

Analysis and Review of Panelists' Feedback

Upon completion of the standard setting evaluation forms, panelists' responses were reviewed. This review did not reveal any anomalies in the standard setting process or indicate any reason that a particular panelist's data should not be included when the final cut scores were calculated. It appeared that all panelists understood the rating task and attended to it appropriately. The results of the standard setting evaluation are included in Appendix H.

Calculation of Recommended Cut Scores

Tables 13a through 16b below present the recommended cuts resulting from the standard setting. There are two tables for each content area: the first shows the results of the Round 3 ratings (i.e., the final recommendations of the standard setting panelists), and the second shows the changes to the Round 3 ratings that were recommended by the policy forum participants.

Table 13a: English Language Arts – Based on Recommended Cuts Following Round 3

Grade	AA 1–	%	AA 2–	%	AA 3–	%	AA 4–	%
	NM		PM		M		MD	
3	0–32	9.2	33–39	12.1	40–44	18.3	45–48	60.4
4	0–32	8.5	33–40	14.7	41–45	22.5	46–48	54.4
5	0–29	6.0	30–38	10.3	39–46	35.3	47–48	48.5
6	0–27	5.4	28–40	19.9	41–45	23.1	46–48	51.6
7	0–21	1.8	22–36	13.1	37–42	17.3	43–48	67.9
8	0–21	2.0	22–34	9.1	35–42	22.4	43–48	66.5
High School	0–28	5.7	29–34	6.1	35–42	20.8	43–48	67.4

Table 13b: English Language Arts – Based on Recommended Cuts Following Policy Forum
 (Highlight shows adjustment recommendations with general consensus by group.)

Grade	AA 1– NM	%	AA 2– PM	%	AA 3– M	%	AA 4– MD	%
3	0–32	9.2	33–39	12.1	40–44	18.3	45–48	60.4
4	0–32	8.5	33–40	14.7	41–45	22.5	46–48	54.4
5	0–29	6.0	30–38	10.3	39–46	35.3	47–48	48.5
6	0–27	5.4	28–40	19.9	41–45	23.1	46–48	51.6
7	0–21	1.8	22–36	13.1	37–42	17.3	43–48	67.9
8	0–21	2.0	22–34	9.1	35–42	22.4	43–48	66.5
High School	0–26	5.0	27–34	6.8	35–42	20.8	43–48	67.4

Table 14a: Mathematics – Based on Recommended Cuts Following Round 3

Grade	AA 1– NM	%	AA 2– PM	%	AA 3– M	%	AA 4– MD	%
3	0–22	2.7	23–34	10.8	35–45	35.1	46–48	51.5
4	0–22	3.2	23–38	17.8	39–46	34.0	47–48	45.1
5	0–24	5.3	25–36	9.0	37–46	35.2	47–48	50.6
6	0–21	1.4	22–33	11.6	34–43	21.7	44–48	65.3
7	0–33	12.2	34–37	7.3	38–46	33.9	47–48	46.5
8	0–31	10.7	32–38	9.0	39–47	42.0	48	38.3
High School	0–19	1.6	20–34	11.4	35–45	35.2	46–48	51.8

Table 14b: Mathematics – Based on Recommended Cuts Following Policy Forum
(Highlight shows adjustment recommendations with general consensus by group.)

Grade	AA 1- NM	%	AA 2- PM	%	AA 3- M	%	AA 4- MD	%
3	0–22	2.7	23–34	10.8	35–45	35.1	46–48	51.5
4	0–22	3.2	23–38	17.8	39–46	34.0	47–48	45.1
5	0–24	5.3	25–36	9.0	37–46	35.2	47–48	50.6
6	0–21	1.4	22–33	11.6	34–44	24.2	45–48	62.9
7	0–31	10.3	32–37	9.2	38–46	33.9	47–48	46.5
8	0–31	10.7	32–38	9.0	39–46	33.8	47–48	46.5
High School	0–19	1.6	20–34	11.4	35–45	35.2	46–48	51.8

Table 15a: Science – Based on Recommended Cuts Following Round 3

Grade	AA 1- NM	%	AA 2- PM	%	AA 3-M	%	AA 4- MD	%
4	0–29	5.6	30–33	3.5	34–43	20.2	44–48	70.7
8	0–28	5.5	29–38	12.8	39–43	15.1	44–48	66.6
High School	0–20	1.9	21–35	10.9	36–43	20.2	44–48	67.0

Table 15b: Science – Based on Recommended Cuts Following Policy Forum
(Highlight shows adjustment recommendations with general consensus by group.)

Grade	AA 1- NM	%	AA 2- PM	%	AA 3-M	%	AA 4- MD	%
4	0–28	5.4	29–33	3.7	34–43	20.2	44–48	70.7
8	0–28	5.5	29–38	12.8	39–43	15.1	44–48	66.6
High School	0–23	2.8	24–35	10.0	36–43	20.2	44–48	67.0

Table 16a: Social Studies – Based on Recommended Cuts Following Round 3

Grade	AA 1– NM	%	AA 2– PM	%	AA 3–M	%	AA 4– MD	%
5	0–33	12.2	34–40	12.3	41–47	32.5	48	42.9
8	0–33	14.4	34–36	3.8	37–47	43.7	48	38.1
High School	0–32	10.8	33–38	7.7	39–46	36.7	47–48	44.8

Table 16b: Social Studies – Based on recommended cuts following Policy Forum
(Highlight shows adjustment recommendations with general consensus by group.)

Grade	AA 1– NM	%	AA 2– PM	%	AA 3–M	%	AA 4– MD	%
5	0–33	12.2	34–40	12.3	41–45	20.5	46–48	55.0
8	0–31	12.4	32–36	5.9	37–45	31.1	46–48	50.7
High School	0–32	10.8	33–38	7.7	39–45	30.1	46–48	51.3

Preparation of This Standard Setting Report

Following final compilation of standard setting results for 2008, Measured Progress prepared this report, which documents the procedures and results of the 2008 standard setting meetings in order to establish performance standards for NYSAA in English language arts, mathematics, science, and social studies.

References

- Angoff, W. H. (1971). Scales, Norms, and Equivalent Scores. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 508-60). Washington, DC: American Council on Education.
- Bay, L. (2000). *1998 NAEP Writing Achievement Levels-Setting Process Performance Profiles Study*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA
- Berk, R. A. (1986). A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests. *Review of Educational Research*, 56, 137-172.
- Cizek, G.J. (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, Gr. J. (1996). Standard Setting Guidelines. *Educational Measurement: issues and Practice*, 15, 13-21.
- Cizek, G.J. & Bunch, M.B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance*. Newbury Park, CA: Sage Publications.
- Hambleton, R. Jaeger, R., Plake, B. & Mills, C. (2000). Setting Performance Standards on Complex Educational Assessments. *Applied Psychological Measurement*, 24 (4), December 2000, 355. 366.
- Hambleton, R. K. (2001) Setting Performance Standards on Educational Assessments and Criteria for Evaluating the Process. In G. J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, N.J.: Erlbaum, 89-116.
- Hambleton, R. K., & Powell, S. (1983). A Framework for Viewing the Process of Standard Setting. *Evaluation & the Health Professions*, 6(1), 3-24.
- Jaeger, R. M. (1995). Setting Performance Standards through Two-Stage Judgmental Policy Capturing. *Applied Measurement in Education*, 8, 15-40.
- Kane, M. (2001). So Much Remains the Same: Conception and Status on Validation in Setting Standards. In G. J. Cizek (ed.), *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum, 53-88.
- Kane, M. T. (1994). Validating the Performance Standards Associated with Passing Scores. *Review of Educational Research*, 64(3), 425-461.
- Kingston, N. M., Kahl, S. R., 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*. Mahwah, NJ: Lawrence Erlbaum, 219-248.

- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). *Standard Setting: A Bookmark Approach*. Paper Presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service.
- Lunz, M. (1995). *Methods of setting criterion standards for performance examinations*. Unpublished manuscript.
- Malehorn, H. (1994). Ten measures better than grading. *The Clearing House*, 67(6), 323-24. [EJ490760].
- Navarrete, C., Wilde, J. Nelson, C., Martinez, R., & Hargett, G. (1990). *Informal Assessment in Educational Evaluation: Implications for Bilingual Education Programs*. Washington, DC: National Clearinghouse for Bilingual Education, 28 pages. [ED337041].
- Reckase, M. D. (2000). A Survey and Evaluation of Recently Developed Procedures for Setting Standards on Educational Tests. In: *Student performance Standards on the National Assessment of Educational progress: Affirmations and Improvement*. Ed. By M. L. Bourqey & Sh. Byrd, Washington: NAEP, pp. 41 . 70.
- Roeber, Ed. (2002). *Setting standards on alternate assessments, Synthesis Report*. NCEO-42. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- SAS Institute (2004). *Base SAS 9.1 Procedures Guide*. Cary, NC: SAS Institute.
- Smith, C. B. (2003). *Alternate Forms of Assessment*. ERIC topical biography and commentary. ERIC Clearinghouse on Reading, English, and Communication, Bloomington, IN.
- Thompson, S. J., & Thurlow, M. L. (2001). *2001 State Special Education Outcomes: A Report on State Activities at the Beginning of a New Decade*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcome.

Appendices

APPENDIX A—MEETING AGENDA



**New York State Alternate Assessment (NYSAA)
Standard Setting
June 9-13, 2008
Crowne Plaza, Albany**

Activity/Presentation	Presenter
Day One	
Sign in and continental breakfast – 8:30 am – 9:00 am	Participants
Welcome, Introductions, and Thank You	Victoria Ferrara New York State Education Department Office of Standards, Assessment and Reporting
Message from the Assistant Commissioner	David Abrams New York State Education Department Office of Standards, Assessment and Reporting
Review Agenda and Materials NYSAA Overview	Kelly Ickes, Measured Progress
Roles of Panelists	Rebecca Walk, Measured Progress
Standard Setting Process	MP Psychometrician, Measured Progress
Break (move to break-out rooms)	Participants
Individual Group Introductions Review Alternate Performance Level Descriptors Performance Level Discussions	Break-out Rooms
Lunch – 12:00 pm – 12:45 pm	General Session Room
Standard Setting Process (1 st grade)	Break-out Rooms
Day Two (Break as needed)	
Continental Breakfast – 8:30 am – 9:00 am	General Session Room
Standard Setting Process (1 st grade)	Break-out Rooms
Lunch – 12:00 pm – 12:45 pm	General Session Room
Standard Setting Process (1 st grade)	Break-out Rooms
Day Three (Break as needed)	
Continental Breakfast – 8:30 am – 9:00 am	General Session Room
Standard Setting Process (1 st grade) Standard Setting Process (2 nd grade) Review Alternate Performance Level Descriptors Performance Level Discussions	Break-out Rooms
Lunch – 12:00 pm – 12:45 pm	General Session Room
Standard Setting Process (2 nd grade)	Break-out Rooms
Day Four (Break as needed)	
Continental Breakfast – 8:30 am – 9:00 am	General Session Room
Standard Setting Process (2 nd grade)	Break-out Rooms
Lunch – 12:00 pm – 12:45 pm	General Session Room
Standard Setting Process (2 nd grade)	Break-out Rooms
Day Five (Break as needed)	
Continental Breakfast – 8:30 am – 9:00 am	General Session Room
Standard Setting Process (2 nd grade)	Break-out Rooms

APPENDIX B—ALTERNATE PERFORMANCE LEVEL DESCRIPTORS

Grade	ELA
3	Reading and listening for literary response and expression AND social interaction
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to and/or listens to simple materials, genre, pictures, characters and/or parts of a story; and/or</i> • <i>Student rarely reads or listens to texts read aloud, sharing texts, taking turns and/or sharing informational texts even with supports; and/or</i> • <i>Student rarely identifies words, conventions, purposes, tone and/or feelings</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to and/or listens with attention to simple materials, genre, pictures, characters and/or parts of a story; and/or</i> • <i>Student inconsistently reads or listens to texts read aloud, sharing texts, taking turns and/or sharing information; and/or</i> • <i>Student inconsistently identifies words, conventions, purposes, tone and/or feelings</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to and/or reads to build comprehension and/or make simple predictions; and/or</i> • <i>Student often engages in the selection of books; and/or</i> • <i>Student often engages in discussion with classmates about texts (i.e., listens to others thoughts and opinions, answers “wh” questions, etc.); and/or</i> • <i>Student often identifies words, conventions, purposes, tone and/or feelings</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently reads with fluency and/or comprehension a variety of texts and is able to identify important details; and/or</i> • <i>Student consistently attends to or reads and/or interacts with classmates; and/or</i> • <i>Student consistently identifies words, conventions, purposes, tone and/or feelings</i>

Grade	ELA
4	Reading for literary response and expression and social interaction AND writing for information and understanding and literary response and expression
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to or reads literary text and/or different genres; and/or</i> • <i>Student rarely responds to questions, thoughts and/or opinions about text; and/or</i> • <i>Student rarely identifies, records, arranges and/or creates pictures, symbols or words</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads literary text and/or different genres; and/or</i> • <i>Student inconsistently responds to questions, thoughts and/or opinions about text; and/or</i> • <i>Student inconsistently identifies, records, arranges and/or creates pictures, symbols or words</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads literary text and/or different genres; and/or</i> • <i>Student often identifies parts of a story and/or motives in text read; and/or</i> • <i>Student often responds to questions, thoughts and/or opinions about text; and/or</i> • <i>Student often identifies, records, arranges and/or creates pictures, symbols or words</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads literary text and/or different genres; and/or</i> • <i>Student consistently identifies parts of a story and/or motives in text read; and/or</i> • <i>Student consistently responds to questions, thoughts and/or opinions about text; and/or</i> • <i>Student consistently identifies, records, arranges and/or creates pictures, symbols or words</i>

Grade	ELA
5	Reading and listening for information and understanding AND literary response and expression
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to or reads to find facts and/or ideas in informational and/or literary texts; and/or</i> • <i>Student rarely recognizes basic literary terms and text features and/or recognizes basic literary devices; and/or</i> • <i>Student rarely follows verbal directions or responds to speaker; and/or</i> • <i>Student rarely recognizes factual statements and main ideas from informational text or listens to take notes</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads to find facts and/or ideas in informational and/or literary texts; and/or</i> • <i>Student inconsistently recognizes or identifies basic literary terms or text features or recognizes basic literary devices; and/or</i> • <i>Student inconsistently follows simple verbal directions or responds to speaker; and/or</i> • <i>Student inconsistently recognizes factual statements and main ideas from informational text or listens to take notes</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads to find facts and/or ideas in informational and/or literary texts; and/or</i> • <i>Student often identifies basic literary terms or text features or recognizes basic literary devices; and/or</i> • <i>Student often follows simple verbal directions or responds to speaker; and/or</i> • <i>Student often recognizes factual statements and main ideas from informational text or listens to take notes</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads to find facts and/or ideas in informational and/or literary texts; and/or</i> • <i>Student consistently identifies basic literary terms or text features or recognizes literary devices(i.e., rhyme, rhythm, personification, simile, etc) and/or</i> • <i>Student consistently follows verbal directions or responds to speaker; and/or</i> • <i>Student consistently recognizes factual statements and main ideas from informational text or listens to take notes</i>

Grade	ELA
6	Reading and writing for information and understanding AND literary response and expression
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely is able to attend to or read texts to collect facts or recognize similarities, differences, predictions and/or conclusions; and/or</i> • <i>Student rarely recognizes characterization within text; and/or</i> • <i>Student rarely uses the writing process to create written text (e.g., pictures, symbols, objects, etc.)</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates [add an]understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads texts to collect facts and/or recognize similarities, differences, predictions and/or conclusions; and/or</i> • <i>Student inconsistently recognizes characterization within text; and/or</i> • <i>Student inconsistently uses the writing process to create written text (e.g., pictures, symbols, objects, etc.)</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads texts to collect facts and/or recognize similarities, differences, predictions and/or conclusions; and/or</i> • <i>Student often recognizes characterization within text; and/or</i> • <i>Student often uses the writing process to create written text (e.g., pictures, symbols, objects, etc.)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads texts to collect facts and/or recognize similarities, differences, predictions and/or conclusions; and/or</i> • <i>Student consistently recognizes characterization within text; and/or</i> • <i>Student consistently uses the writing process to create written text (e.g., pictures, symbols, objects, etc.)</i>

Grade	ELA
7	Reading and listening for information and understanding AND literary response and expression
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to or reads to recognize the distinction between fact and opinion and/or predictions, comparisons, or conclusions based on text; and/or</i> • <i>Student rarely identifies literary device(s) (i.e., rhyme, rhythm, poetry, similes, metaphors, main ideas, theme, etc.) and/or literary element(s) (i.e., character, plot, setting, etc.); and/or</i> • <i>Student rarely demonstrates the ability to listen to a speaker and/or the ability to identify essential information</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads to recognize the distinction between fact and opinion and/or predictions, comparisons, or conclusions based on text; and/or</i> • <i>Student inconsistently identifies literary device(s) (i.e., rhyme, rhythm, poetry, similes, metaphors, main ideas, theme, etc.) and/or literary element(s) (i.e., character, plot, setting, etc.); and/or</i> • <i>Student inconsistently demonstrates the ability to listen to a speaker and/or the ability to identify essential information</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads to recognize the distinction between fact and opinion and/or predictions, comparisons, or conclusions based on text; and/or</i> • <i>Student often identifies literary device(s) (i.e., rhyme, rhythm, poetry, similes, metaphors, main ideas, theme, etc.) and/or literary element(s) (i.e., character, plot, setting, etc.); and/or</i> • <i>Student often demonstrates the ability to listen to a speaker and/or the ability to identify essential information</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads to recognize the distinction between fact and opinion and/or predictions, comparisons, or conclusions based on text; and/or</i> • <i>Student consistently identifies literary device(s) (i.e., rhyme, rhythm, poetry, similes, metaphors, main ideas, theme, etc.) and/or literary element(s) (i.e., character, plot, setting, etc.); and/or</i> • <i>Student consistently demonstrates the ability to listen to a speaker and/or the ability to identify essential information for note-taking</i>

Grade	ELA
8	Reading and writing for information and understanding AND critical analysis and evaluation
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to or reads text to collect facts, compare or contrast subjects and/or draw conclusions based on text; and/or</i> • <i>Student rarely demonstrates the use of problem solving skills and the ability to think analytically (i.e., identify perspective, opinion and validity of information, etc.) or make predictions or form an opinion; and/or</i> • <i>Student rarely demonstrates the ability to record, summarize or use persuasive writing (e.g., pictures, symbols, objects, etc.)</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads text to collect facts, compare or contrast subjects and/or draw conclusions based on text; and/or</i> • <i>Student inconsistently demonstrates the use of problem solving skills and the ability to think analytically (i.e., identify perspective, opinion and validity of information, etc.) or make predictions or form an opinion; and/or</i> • <i>Student inconsistently demonstrates the ability to record, summarize or use persuasive writing (e.g., pictures, symbols, objects, etc.)</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads text to collect facts, compare and/or contrast subjects and/or draw conclusions based on text; and/or</i> • <i>Student often demonstrates the use of problem solving skills and the ability to think analytically (i.e., identify perspective, opinion and validity of information, etc.) or make predictions or form an opinion; and/or</i> • <i>Student often demonstrates ability to record, summarize or use persuasive writing (e.g., pictures, symbols, objects, etc.)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads text to collect facts, compare and/or contrast subjects and/or draw conclusions based on text; and/or</i> • <i>Student consistently demonstrates the use of problem solving skills and the ability to think analytically (i.e., identify perspective, opinion and validity of information, etc.) or make predictions or form an opinion; and/or</i> • <i>Student consistently demonstrates ability to record, summarize or use persuasive writing (e.g., pictures, symbols, objects, etc.) with personal voice</i>

Grade	ELA
HS	Reading and writing for information and understanding AND critical analysis and evaluation
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely attends to or reads to gather and/or interpret a variety of information from different sources and/or locations; and/or</i> • <i>Student rarely attends to or reads to answer comprehension questions and/or summarize informational text; and/or</i> • <i>Student rarely creates or writes (e.g., pictures, symbols, objects, etc.) a basic persuasive, expository or descriptive piece on a topic(s)</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently attends to or reads to gather and/or interpret a variety of information from different sources and/or locations; and/or</i> • <i>Student inconsistently attends to or reads to answer comprehension questions and/or summarize informational text; and/or</i> • <i>Student inconsistently creates or writes (e.g., pictures, symbols, objects, etc.) a basic persuasive, expository or descriptive piece on a topic(s)</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often attends to or reads to gather and/or interpret a variety of information from different sources and/or locations; and/or</i> • <i>Student often attends to or reads to answer comprehension questions and/or summarize informational text; and/or</i> • <i>Student often creates or writes (e.g., pictures, symbols, objects, etc.) a basic persuasive, expository or descriptive piece on a topic(s)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently attends to or reads to gather and/or interpret a variety of information from different sources and/or locations; and/or</i> • <i>Student consistently attends to or reads to answer comprehension questions and/or summarize informational text; and/or</i> • <i>Student consistently creates or writes (e.g., pictures, symbols, objects, etc.) a basic persuasive, expository or descriptive piece on a topic(s)</i>

Grade	Mathematics
3	Number Sense and operations, including number systems and operations AND Measurement including units of measurement and units/estimation
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses number sense to recognize numbers, compare quantities or use operations to solve problems using basic addition; and/or</i> • <i>Student rarely recognizes and/or uses measurement tools or tells time</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses number sense to identify numbers, compare quantities or use operations to solve problems using basic addition or subtraction; and/or</i> • <i>Student inconsistently recognizes and uses measurement tools, tells time or recognizes money</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses number sense to identify numbers, compare quantities or use operations to solve problems using basic addition or subtraction; and/or</i> • <i>Student often recognizes and/or uses measurement tools, tells time or recognizes money</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses number sense to identify numbers and/or fractions, compare quantities or use operations to solve problems using basic calculations; and/or</i> • <i>Student consistently recognizes and/or uses measurement tools, tells time or recognizes money and its value or denomination</i>

Grade	Mathematics
4	Number Sense and operations, including number systems and operations AND Measurement including units of measurement and units/estimation
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses number sense to recognize numbers, compare quantities or use operations to solve problems using basic addition and/or subtraction; and/or</i> • <i>Student rarely recognizes and/or uses measurement tools, tells time or recognizes money including its value or denomination</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses number sense strategies to identify numbers and/or decimals, compare quantities of whole objects or use operations to solve problems using basic addition and/or subtraction; and/or</i> • <i>Student inconsistently recognizes and uses measurement tools, tells time or recognizes money including its value or denomination</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses number sense strategies to identify numbers to count, group, compare, and/or order whole numbers; or</i> • <i>Student often recognizes fractions or decimals; and/or</i> • <i>Student often recognizes and uses measurement tools, tells time or recognizes money including its value or denomination</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses strategies to identify numbers, fractions, and/or decimals to compare quantities using greater than, less than, or equal to, or to count, group and/or order numbers; and/or</i> • <i>Student consistently uses operation strategies to perform whole-number computations (i.e., addition, subtraction, etc.); and/or</i> • <i>Student consistently recognizes and uses measurement tools, tells time or identifies money and/or equivalent coin combinations</i>

Grade	Mathematics
5	Number Sense and Operations including number systems and operations AND Geometry including geometric relationships and coordinate geometry
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses number sense to recognize numbers and value, compare quantities using greater than, less than, or equal to or uses operations to solve problems using basic addition and/or subtraction; and/or</i> • <i>Student rarely recognizes shapes or locates a position on a number line</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses number sense strategies to identify numbers and value, identify fractions and/or decimals, compare quantities using greater than, less than, or equal to or recognize a whole and its parts; and/or</i> • <i>Student inconsistently uses operations to solve problems using basic addition and/or subtraction; and/or</i> • <i>Student inconsistently recognizes and sorts shapes or locates positions on a number line</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses number sense strategies to identify numbers and value or to count, group, compare, and/or order whole numbers</i> • <i>Student often recognizes fractions or decimals or matches fractions to concrete representations; and/or</i> • <i>Student often uses operations to solve problems using basic computation; and/or</i> • <i>Student often identifies and/or sorts shapes or plots a point on a number line or grid</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student's performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses strategies to identify numbers and value, identify fractions, and/or decimals, compare quantities using greater than, less than, or equal to, or to count, group and/or order numbers; and/or</i> • <i>Student consistently uses operation strategies to perform whole-number computations or fractions (i.e., addition, subtraction, etc.); and/or</i> • <i>Student consistently identifies and/or classifies shapes by properties or plots a point on a number line or grid when given its coordinates</i>

Grade	Mathematics
6	Number Sense and Operations including number systems and operations AND Algebra including variables and expressions and equations and inequalities
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses number sense or operation strategies to identify numbers to count, group, compare, and/or order whole numbers; and/or</i> • <i>Student rarely recognizes fractions or decimals or matches fractions to concrete representations; and/or</i> • <i>Student rarely uses concrete objects to solve simple algebraic equations or recognizes place value related to money</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses number sense or operation strategies to identify numbers to count, group, compare, and/or order whole numbers; and/ or</i> • <i>Student inconsistently recognizes fractions or decimals or matches fractions to concrete representations; and/or</i> • <i>Student inconsistently uses concrete objects to solve simple algebraic equations or recognizes place value or rounds related to money</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses number sense or operation strategies to identify numbers to count, group, compare, and/or order whole numbers; and/or</i> • <i>Student often recognizes fractions or decimals or uses a strategy to perform basic computations with fractions; and/or</i> • <i>Student often uses a strategy to solve simple algebraic equations, recognizes place value or rounds to 100 (usually related to money)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses number sense or operation strategies to identify numbers to count, group, compare, and/or order whole numbers; and or</i> • <i>Student consistently recognizes fractions or decimals or uses a strategy to perform basic computations with fractions; and/or</i> • <i>Student consistently uses a strategy to solve simple algebraic equations, recognizes place value or rounds to 100</i>

Grade	Mathematics
7	Number Sense and Operations including number theory and operations AND Statistics and Probability including collection of data, organization and display of data and analysis of data
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses number theory to recognize numbers or factors with concrete objects and operations to solve problems using basic computation; and/or</i> • <i>Student rarely gathers and/or recognizes relevant data</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses number theory to recognize numbers or factors with concrete objects and operations to solve problems using basic computation; and/or</i> • <i>Student inconsistently locates integers on a number line; and/or</i> • <i>Student inconsistently gathers and/or recognizes relevant data</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses number theory to recognize numbers or factors to solve problems of basic computation (i.e., addition, subtraction, multiplication, etc.) using one or more strategies; and/or</i> • <i>Student often locates and/or adds integers on a number line; and/or</i> • <i>Student often gathers and/or organizes relevant data</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses number theory to recognize numbers or factors to solve problems of basic computation (i.e., addition, subtraction, multiplication, etc.) using one or more strategies; and/or</i> • <i>Student consistently adds or subtracts integers using a number line; and/or</i> • <i>Student consistently gathers and/or interprets relevant data</i>

Grade	Mathematics
8	Geometry including geometric relationships and transformational geometry AND Algebra including variables and expressions and patterns, relations and functions
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely recognizes shapes and/or uses trial and error to demonstrate transformations (i.e., translations, etc.); and/or</i> • <i>Student rarely solves simple algebraic equations with or without concrete objects and/or recognizes numbers but not values, and/or compare quantities with concrete objects using terms greater than, less than, or equal to; and/or</i> • <i>Student rarely duplicates a pattern given a model</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently recognizes shapes or angles and/or uses trial and error to demonstrate transformations (i.e., translations, reflections, rotations, etc.); and/or</i> • <i>Student inconsistently solves simple algebraic equations with or without concrete objects and/or recognizes numbers and values, and/or compares quantities with concrete objects using terms greater than, less than, or equal to; and/or</i> • <i>Student inconsistently duplicates or extends a pattern given a model</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often identifies shapes or congruent angles and/or demonstrates transformations (i.e., translations, reflections, rotations, dilations, etc.) using concrete objects; and/or</i> • <i>Student often translates or solves simple algebraic equations with or without concrete objects and/or compares quantities of objects or numbers using terms greater than, less than, equal to, or not equal; and/or</i> • <i>Student often extends a growing or repeating pattern or creates a pattern</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student's performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently identifies shapes, types of angles or angle pairs (i.e., vertical, complementary, etc.) and/or demonstrates the use of transformations (i.e., translations, reflections, rotations, dilations, etc.) using purposeful movement or action; and/or</i> • <i>Student consistently translates or solves simple algebraic equations with or without concrete objects and/or compares quantity expressions using equal to, or not equal; and/or</i> • <i>Student consistently extends a growing or repeating pattern or creates a pattern</i>

Grade	Mathematics
HS	Algebra including variables and expressions and equations and inequalities AND Statistics and Probability including organization and display of data and analysis of data
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support(i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses concrete objects, symbols, or visuals to translate or to solve simple algebraic equations; and/or</i> • <i>Student rarely compares quantities with concrete objects using terms greater than, less than, or equal to; and/or</i> • <i>Student rarely describes or extends a pattern given a model; and/or</i> • <i>Student rarely gathers and/or recognizes data</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses concrete objects, symbols, or visuals to translate or to solve simple algebraic equations; and/or</i> • <i>Student inconsistently identifies a correct numerical expression and/or compares quantities of objects or numbers using terms greater than, less than, or equal to; and/or</i> • <i>Student inconsistently describes or extends a numeric or shape pattern given a model; and/or</i> • <i>Student inconsistently gathers, recognizes, and/or interprets data</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often translates or solves simple algebraic equations; and/or</i> • <i>Student often identifies a correct numerical expression and/or compares quantities of objects or numbers using terms greater than, less than, or equal to; and/or</i> • <i>Student often describes or continues a numeric or shape pattern given a model; and/or</i> • <i>Student often gathers, recognizes, and/or interprets data</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student's performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently translates or solves simple algebraic equations; and/or</i> • <i>Student consistently identifies a correct numerical expression and/or compares quantities of objects or numbers using terms greater than, less than, or equal to; and/or</i> • <i>Student consistently describes or continues a numeric or shape pattern; and/or</i> • <i>Student consistently gathers, recognizes, and/or interprets data</i>

Grade	Science
4	Scientific Inquiry AND Living Environment and Physical Setting/Earth Science
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely makes observations, interacts with objects, or recognizes similarities and differences, and/or outcomes of a procedure; and/or</i> • <i>Student rarely recognizes tools, objects, and/or procedures that are associated with scientific inquiry; and/or</i> • <i>Student rarely understands or identifies aspects of the living environment (e.g., structures such as wings, eyes, roots, leaves, survival adaptations, etc.) and/or physical setting (e.g., water cycle, natural events, erosion, etc.) of the Earth</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently makes observations, interacts with objects, or recognizes similarities and differences, and/or outcomes of a procedure; and/or</i> • <i>Student inconsistently recognizes tools, objects, and/or procedures that are associated with scientific inquiry; and/or</i> • <i>Student inconsistently understands or identifies aspects of the living environment (e.g., structures such as wings, eyes, roots, leaves, survival adaptations, etc.) or physical setting (e.g., water cycle, natural events, erosion, etc.) of the Earth</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often makes observations, interacts with objects, or recognizes similarities and differences, and/or outcomes of a procedure; and/or</i> • <i>Student often recognizes tools, objects, and/or procedures that are associated with scientific inquiry; and/or</i> • <i>Student often understands or identifies aspects of the living environment (e.g., structures such as wings, eyes, roots, leaves, survival adaptations, etc.) or physical setting (e.g., water cycle, natural events, erosion, etc.) of the Earth</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently makes observations, interacts with objects, or recognizes similarities and difference, and outcome of a procedures; and/or</i> • <i>Student consistently recognizes tools, objects, and/or procedures that are associated with scientific inquiry; and/or</i> • <i>Student consistently understands and identifies aspects of the living environment (e.g., structures such as wings, eyes, roots, leaves, survival adaptations, etc.) or physical setting (e.g., water cycle, natural events, erosion, etc.) of the Earth</i>

Grade	Science
8	Scientific Inquiry AND Living Environment and Physical Setting/Earth Science
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum supports (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely uses scientific inquiry to conduct scientific investigations, record and/or organize data, and/or draw conclusions; and/or</i> • <i>Student rarely recognizes cause and effect and/or compares results in tables, graphs, etc; and/or</i> • <i>Student rarely recognizes the aspects of organisms that sustain life (i.e., survival adaptations, structures to obtain energy and nutrients, maintaining a balanced system, etc.) and/or that matter has observable characteristics (i.e., properties of matter, physical change, etc.)</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently uses scientific inquiry to conduct scientific investigations, record and/or organize data, and/or draw conclusions; and/or</i> • <i>Student inconsistently recognizes cause and effect, and/or compares results in tables, graphs, etc; and/or</i> • <i>Student inconsistently understands or identifies the aspects of organisms that sustain life (i.e., survival adaptations, structures to obtain energy and nutrients, maintaining a balanced system, etc.) and/or that matter has observable characteristics (i.e., properties of matter, physical change, etc.)</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often uses scientific inquiry to conduct scientific investigations, record and/or organize data, and/or draw conclusions; and/or</i> • <i>Student often recognizes cause and effect, and/or compares results in tables, graphs, etc; and/or</i> • <i>Student often understands or identifies the aspects of organisms that sustain life (i.e., survival adaptations, structures to obtain energy and nutrients, maintaining a balanced system, etc.) and/or that matter has observable characteristics (i.e., properties of matter, physical change, etc.)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently uses scientific inquiry to conduct scientific investigations, record and/or organize data, and/or draw conclusions; and/or</i> • <i>Student consistently recognizes cause and effect, and/or compares results in tables, graphs, etc; and/or</i> • <i>Student consistently understands and identifies the aspects of organisms that sustain life (i.e., survival adaptations, structures to obtain energy and nutrients, maintaining a balanced system, etc.) and/or that matter has observable characteristics (i.e., properties of matter, physical change, etc.)</i>

Grade	Science
HS	Living Environment AND Physical Setting/Earth Science
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely recognizes that humans have structures called organs; and/or</i> • <i>Student rarely recognizes or demonstrates that human decisions and activities have an impact on the environment; and/or</i> • <i>Student rarely recognizes that all phenomena (i.e., water cycle, planet movement, etc.) in the universe occur as a result of the principles of either relative motion and perspective (i.e., size, distance, time) or interactions among air, water, and land (i.e., weather conditions, seasons, heat source, etc.)</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently recognizes the similarities and differences between living and nonliving things</i> • <i>Student inconsistently recognizes that humans have structures called organs; and/or</i> • <i>Student inconsistently recognizes or demonstrates that human decisions and activities have an impact on the environment; and/or</i> • <i>Student inconsistently recognizes that all phenomena (i.e., water cycle, planet movement, etc.) in the universe occur as a result of the principles of either relative motion and perspective (i.e., size, distance, time) or interactions among air, water, and land (i.e., weather conditions, seasons, heat source, etc.)</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student's performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often identifies the similarities and differences between living and nonliving things</i> • <i>Student often recognizes that humans have structures called organs; and/or</i> • <i>Student often identifies or demonstrates that human decisions and activities have an impact on the environment; and/or</i> • <i>Student often identifies that all phenomena (i.e., water cycle, planet movement, etc.) in the universe occur as a result of the principles of either relative motion and perspective (i.e., size, distance, time) or interactions among air, water, and land (i.e., weather conditions, seasons, heat source, etc.)</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student's performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently identifies the similarities and differences between living and nonliving things</i> • <i>Student consistently recognizes that humans have structures called organs; and/or</i> • <i>Student consistently identifies or demonstrates that human decisions and activities have an impact on the environment; and/or</i> • <i>Student consistently identifies that all phenomena (i.e., water cycle, planet movement, etc.) in the universe occur as a result of the principles of either relative motion and perspective (i.e., size, distance, time) or interactions among air, water, and land (i.e., weather conditions, seasons, heat source, etc.)</i>

Grade	Social Studies
5	US and NYS History including colonial life and revolutionary war in NYS and industrial growth and expansion in NYS AND Civics, Citizenship and Government including government of world communities and local, state and national government
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely demonstrates an understanding of colonial life in NYS and US History and elements related to industrialization such as communication, transportation, etc.; and/or</i> • <i>Student rarely demonstrates knowledge of geographic concepts; and/or</i> • <i>Student rarely identifies holidays and/or symbols associated with the United States and/or acceptable and unacceptable behavior in school, home and community based on rules, laws or proper conduct for good citizenship</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently demonstrates an understanding of colonial life in NYS and US History and elements related to industrialization such as communication, transportation, etc.; and/or</i> • <i>Student inconsistently demonstrates knowledge of geographic concepts such as location of home or classroom; and/or</i> • <i>Student inconsistently identifies holidays and/or symbols associated with the United States and/or acceptable and unacceptable behavior in school, home and community based on rules, laws or proper conduct for good citizenship</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often demonstrates an understanding of the elements of colonial life in NYS and US History and changes related to industrialization such as changes in immigration communication, transportation, etc.; and/or</i> • <i>Student often demonstrates knowledge of geographic concepts such as location of state or county of residence; and/or</i> • <i>Student often identifies national holidays and/or symbols associated with the United States and foreign countries and/or acceptable and unacceptable behavior in school, home and community based on rules, laws or proper conduct for good citizenship</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently identifies and compares elements of colonial life in NYS and US History with modern society and changes or impact related to industrialization such as changes in immigration, communication, transportation, etc.; and/or</i> • <i>Student consistently demonstrates knowledge of geographic concepts such as location and/or natural features of state or county of residence; and/or</i> • <i>Student consistently identifies national holidays and/or symbols associated with the United States and foreign countries and/or the process for creating rules and laws or the rights and responsibilities of citizens and their impacts</i>

Grade	Social Studies
8	US and NYS History including industrial society and between the wars AND Civics, Citizenship and Government including experiment in Government and WWII to the present
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely demonstrates recognition of facts related to early American culture and the idea that groups of people came to the United States from different places (i.e., cultural contributions, geographic locations, etc.); and/or</i> • <i>Student rarely recognizes famous people associated with American culture in the 1920s; and/or</i> • <i>Student rarely recognizes the idea that citizens live under governments and/or recognizes government structures (i.e., branches of government, constitutions, civil rights, etc.) and/or the process for creating rules and laws or the rights and responsibilities of citizens and their impacts</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently demonstrates recognition of facts related to early American culture and the idea that groups of people came to the United States from different places (i.e., cultural contributions, geographic locations, etc.); and/or</i> • <i>Student inconsistently recognizes famous people associated with American culture in the 1920s; and/or</i> • <i>Student inconsistently recognizes the idea that citizens live under governments and/or recognizes government structures or events (i.e., branches of government, constitutions, civil rights, etc.) and/or the process for creating rules and laws or rights and responsibilities of citizens and their impacts</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often identifies and explores facts related to early American culture and the idea that groups of people came to the United States from different places (i.e., cultural contributions, geographic locations, roots, immigrant experience, etc.); and/or</i> • <i>Student often recognizes famous people associated with American culture in the 1920s and their contributions; and/or</i> • <i>Student often demonstrates and identifies the idea that citizens live under governments and/or recognizes government structure or events and functions (i.e., branches of government, constitutions, civil rights, etc.) and/or the process for creating rules and laws or the rights and responsibilities of citizens and their impacts</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently identifies and explores facts related to early American culture changes and the idea that immigrant groups influenced American life (i.e., cultural contributions, geographic locations, challenges faced, etc.); and/or</i> • <i>Student consistently recognizes famous people associated with American culture in the 1920s and their contributions; and/or</i> • <i>Student consistently demonstrates and identifies the idea that citizens live under governments and/or identifies government structures or events and functions (i.e., branches of government, constitutions, civil rights, etc.) and/or the process for creating rules and laws or the rights and responsibilities of citizens and their impacts</i>

Grade	Social Studies
HS	US History including constitutional foundations and world in uncertain times: 1980-present AND Global History including age of revolution and global connections and interactions
Not Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is rarely accurate • Student rarely demonstrates an understanding of AGLIs • Student rarely engages in tasks even with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student rarely recognizes classroom and school rules, rights and responsibilities associated with being a citizen, and/or rights guaranteed by the Bill of Rights; and/or</i> • <i>Student rarely recognizes that governments create laws and/or recognizes recent presidents or their actions; and/or</i> • <i>Student rarely demonstrates or recognizes elements of the Industrial Revolution, geographic locations of countries near the United States, and/or needs and wants of people living in different parts of the world</i>
Partially Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is inconsistently accurate • Student inconsistently demonstrates an understanding of AGLIs • Student engages in tasks with maximum support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student inconsistently identifies classroom and school rules, recognizes rights and responsibilities associated with being a citizen, and/or rights guaranteed by the Bill of Rights and/or recognize the Constitution; and/or</i> • <i>Student inconsistently recognizes that governments create laws and/or recognizes recent presidents and their actions; and/or</i> • <i>Student inconsistently demonstrates or recognizes elements of the Industrial Revolution, geographic locations of countries near the United States, and/or needs and wants of people living in different parts of the world</i>
Meeting Learning Standards	<ul style="list-style-type: none"> • Student’s performance is often accurate • Student often demonstrates an understanding of AGLIs • Student actively engages in tasks with occasional support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student often identifies classroom and school rules, recognizes rights and responsibilities associated with being a citizen, and/or identifies the Constitution and the Bill of Rights as the basis for citizen rights in the United States; and/ or</i> • <i>Student often recognizes that governments create laws and/or constitutional principles and/or recognizes current events associated with recent presidents; and/or</i> • <i>Student often demonstrates or recognizes elements of the Industrial Revolution and urbanization, geographic locations of countries other than the United States, and/or identify differences in lifestyle among people living in different parts of the world</i>
Meeting Learning Standards with Distinction	<ul style="list-style-type: none"> • Student’s performance is consistently accurate • Student consistently demonstrates an understanding of AGLIs • Student actively engages in tasks with minimal support (i.e., cues and/or prompts beyond scope of IEP accommodations) • <i>Student consistently identifies classroom and school rules, recognizes rights and responsibilities associated with being a citizen, and/or identifies how the Constitution and the Bill of Rights protect and/or secures the rights of citizens; and/or</i> • <i>Student consistently recognizes how governments create laws and/or constitutional principles and/or recognizes domestic and foreign current events; and/or</i> • <i>Student consistently demonstrates or recognizes elements of the Industrial Revolution and urbanization, geographic locations of developing countries, and/or compare differences in culture and lifestyle among people living in different parts of the world</i>

APPENDIX C—OPENING SESSION POWERPOINT PRESENTATIONS

Slide 1

**New York State
Alternate Assessment**

**ELA, Mathematics, Science, and
Social Studies**

**Standard Setting
June 9-13, 2008**

 University of the
State of New York
State Education
Department

 Measured
Progress

Slide 2

Housekeeping Items

- Please turn cell phones to vibrate
- Please leave room if you need to take a call or return a text message
- Feel free to move about the room and take care of personal needs
- Take advantage of the refreshments while they last

Slide 3

Introductions

- NYS Education Department**
 - Office of Standards, Assessment and Reporting (OSAR)
 - David Abrams – Assistant Commissioner
 - Steven Katz - Director
 - Candy Shyer – Bureau Chief
 - Victoria Ferrara – Program Manager
 - Jacqueline Harnett – Assistant in the Education of Children with Disabilities
 - Office of Vocational and Education Services for Individuals with Disabilities (VESID)

Slide 4

Introductions (con't.)

**Alternate Assessment Contractor:
Measured Progress**

Facilitators	
Rebecca Walk	Alex Ma
Pam Green	Chris Paris
Kelly Ickes	Brook Loch
Alicia Cuttle	Sharon Houle
Barbara Haig	Pat Dubois
Richard Tracey	Karen Orlando

Slide 5

Introductions (con't.)

Research and Analysis

- Liz Burton, Psychometrician
- Luz Bay, Psychometrician
- Woreen-Ann Bogle, Data Analyst

Logistics Assistance

- Michelle Couture
- Kevin Froton
- Sharman Lyons

Slide 6

Process - Agenda

- Welcome, Introductions, and Materials
- Message from the Assistant Commissioner
- NYSAA Overview
- Role of the Panelists
- Standard Setting Process
- Break Out Groups

Slide 7

Purpose

- Recommend the cut scores for reporting Alternate Performance Levels
- Provide suggestions for revision of the Alternate Performance Level Descriptors
- Needs to be completed due to the assessment design changes implemented in 2006-07

Slide 8

Thank You For Your Participation

New York State Alternate Assessment Standard Setting

NYSAA Overview Training
June 9, 2008



University of the
State of New York
State Education
Department



NYSAA Overview

- Datafolio style assessment
 - Students eligible for NYSAA must be assessed on grade level content consistent with the student's chronological age
 - 2:3:2 – the Magic Formula
 - 2 AGLIs for each content area
 - 3 dates of student performance data per AGLI
 - 2 pieces of verifying evidence per AGLI
-

Who takes the Alternate Assessment?



- Meet age, eligibility and participation criteria
 - Students with disabilities who
 - have severe cognitive abilities and significant deficits in communication/language and significant deficits in adaptive behavior and
 - require a highly specialized educational program; and
 - require educational support systems, such as assistive technology, personal care services, health/medical services or behavioral intervention
-

What is an Alternate Assessment?

- An assessment based on alternate academic achievement standards linked to grade-level content. In NYS, this is a datafolio assessment.
 - All students must be included in state's accountability (NCLB, IDEA)
 - Students with severe cognitive disabilities
 - Only a very small percentage of students with disabilities are eligible for an alternate assessment
 - Unable to participate in other state assessments with or without accommodations
-

NYSAA Frameworks

- NYS Learning Standards
 - Core Curriculum in English language Arts, Mathematics, Science, and Social Studies
 - NYSAA Frameworks
 - Grade Level Indicators
 - Essences
 - Alternate Grade Level Indicators (AGLIs)
 - Assessment Tasks
-

Alignment with Grade Level Performance Expectations

- AGLIs are aligned with grade level performance expectations within the core curriculums for
 - grades 3-8 and high school in ELA and mathematics
 - the elementary, middle and high school levels for science and social studies
-

NYSAA – Alignment (con't.)

- AGLIs
 - Are based on essences
 - Reflect different levels of complexity
 - Are what the student is assessed against
 - Are determined by the teacher and instructional team
 - Assessment Tasks
 - Describe an action
 - Are observable
 - Are aligned to AGLIs
 - Are individualized for each student
-

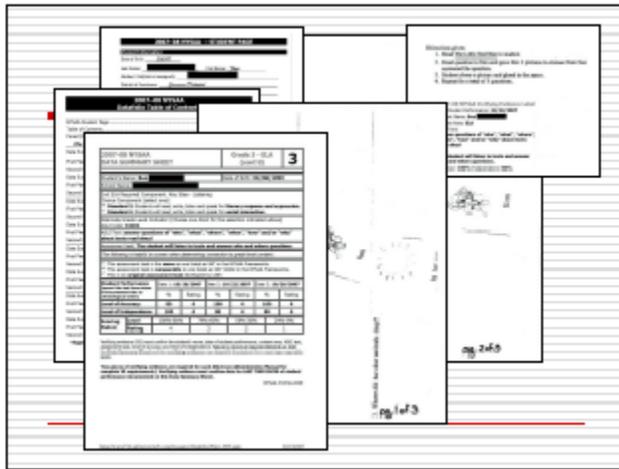
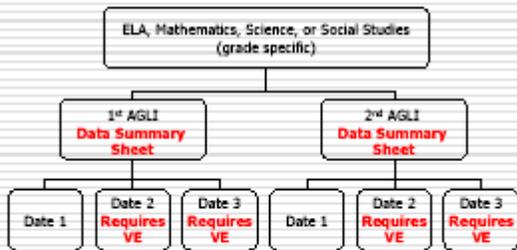
What does the assessment look like?

- Datafolio-style assessment
 - Collection of student assessment data
 - Input from the student, family, and instructional team
 - Includes student performance data
 - Student work products
 - Photographs
 - Video tape and/or Audio tape clips
 - Data Collection Sheets
-

Datafolio Components

- **Data Summary Sheet (DSS)**
 - Summarizes student performance data
 - One for each assessed AGLI
 - **Verifying Evidence (VE)**
 - Confirms student performance data
 - Two pieces of VE are required for each AGLI
-

NYSAA Assessment Design



2007-08 NYSAA DATA SUMMARY SHEET		Grade 3 - ELA (cont'd)		3		
Student's Name: Ben [REDACTED]			Date of Birth: 04/06/1999			
School Name: [REDACTED]						
2nd ELA Required Component: Key Idea - Listening Choice Component (select one): <input checked="" type="radio"/> Standard 2: Students will read, write, listen and speak for literary response and expression. <input type="radio"/> Standard 4: Students will read, write, listen and speak for social interaction.						
Alternate Grade Level Indicator (Choose one AGLI for the selection indicated above) AGLI Code: 32202						
AGLI Text: answer questions of "who", "what", "where", "when", "how" and/or "why" about texts read aloud						
Assessment task: The student will listen to texts and answer who and where questions.						
<i>The following is helpful to scorers when determining connection to grade level content:</i>						
<input type="radio"/> This assessment task is the same as one listed as SAT in the NYSAA Frameworks. <input checked="" type="radio"/> This assessment task is comparable to one listed as SAT 32202 in the NYSAA Frameworks. <input type="radio"/> This is an original assessment task developed by LSET.						
Student Performance (record the last three dates of documented data in chronological order)	Date 1: 10/18/2007		Date 2: 10/22/2007		Date 3: 10/26/2007	
	%	Rating	%	Rating	%	Rating
	80	4	100	4	100	4
Level of Accuracy	80	4	100	4	100	4
Level of Independence	100	4	80	4	80	4
Scoring Rubric	Level	100%-80%	79%-60%	59%-30%	29%-0%	
	Rating	4	3	2	1	

Verifying evidence (VE) must confirm the student's name, date of student performance, content area, AGLI text, assessment task, level of accuracy and level of independence. Failure to record all required elements on both the **Data Summary Sheet** and the **verifying evidence** may disqualify the student from receiving a reportable score.

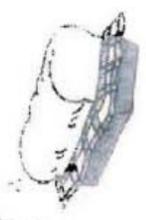
Two pieces of verifying evidence are required for each AGLI (see Administration Manual for complete VE requirements). Verifying evidence must confirm data for LAST TWO DATES of student performance documented on this Data Summary Sheet.

NYSAA ProFile 2008

Don't know
5/5
Verbal prompt

The Little Red Hen

1. What did the Little Red Hen make?



bread

2. Who did all of the work?



hen

3. Where did the other animals sleep?

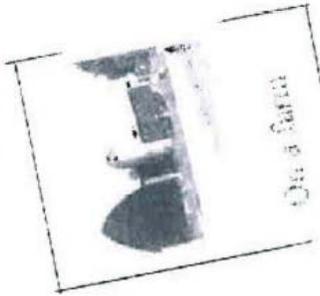


In the ems

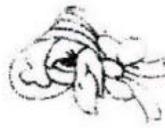
pg. 1 of 3

The Little Red Hen

4. Where did the Little Red Hen live?



5. Who ate the bread?



Hen

pg. 2 of 3

Directions given:

1. Read The Little Red Hen to student
2. Read question to him and gave him 3 pictures to choose from that answered the question.
3. Student chose a picture and glued in the space.
4. Repeat for a total of 5 questions.

2007-08 NYSAA Verifying Evidence Label

Date Student Performance: **10/22/2007**

Student Name: **Ben** [REDACTED]

Content Area: **ELA**

AGLI Text:

**answer questions of "who", "what", "where",
"when", "how" and/or "why" about texts
read aloud**

Task:

**The student will listen to texts and answer
who and where questions.**

Accuracy: **100%** Independence: **80%**

pg. 3 of 3

Example of Photographs

Allison [REDACTED] 11/02/07 ELA AGLI CODE 31102



1) Allison listens to the speaker ask her a question about *The Halloween Story*.



2) Allison responds to the question by touching the Yes switch.



3) Allison responds to another question by touching the Yes switch.

Example of Data Collection Sheet

2007-08 NYSAA Data Collection Sheet for Multi-Step Task										
Student Name: Jeremy [REDACTED]					Content Area: Mathematics					
AGLI text: when given a repeating number and/or shape pattern, extend the pattern										
Assessment task: Jeremy will extend a pattern when given a repeating shape pattern by selecting the correct shape from a choice of 3.										
ACCURACY KEY: (+) Correct Response (-) Inaccurate/No Response INDEPENDENCE KEY: (+) Independent (-) Prompted										
Step Description:	Date 11/27/2007		Date 11/28/2007		Date 11/29/2007		Date No Date		Date No Date	
	Acc +/-	Ind +/-	Acc +/-	Ind +/-	Acc +/-	Ind +/-	Acc +/-	Ind +/-	Acc +/-	Ind +/-
1st pattern	+	+	+	+	+	+				
2nd pattern	+	+	-	+	+	+				
3rd pattern	-	+	-	+	-	+				
Total +'s	2	3	1	3	2	3				
Total Steps	3	3	3	3	3	3				
Fraction	2/3	3/3	1/3	3/3	2/3	3/3				
Percent (%)	66	100	33	100	66	100				
Staff Initials (Req'd)	CC		AP		AP					
Setting	C		C		C					
STAFF KEY (REQUIRED) Name: Carole [REDACTED] Initials: CC					SETTING KEY					
Name: Anne [REDACTED] Initials: AP					(C) Classroom (F) Cafeteria (L) Library (G) Gym					
Name: Initials:					(O) Other specify:					
NOTE: Data Collection Sheet cannot stand alone, supporting evidence is required. Complete in full, including staff initials for each date.										

NYSAA Scoring Training 2007-08

Qualifier #1 page 3

Scoring Dimensions – The Rubric

2007-08 NYSAA SCORING RUBRIC																						
For each content area of each grade, two NYSAA tests are prepared on at least three dates within the administration period. Student scores on the two Required Components for each grade and content area determine the NYSED Performance in Appendix G.																						
Content Area	Grade 3	Grade 4	Grade 5	Grade 6																		
ELA	<ul style="list-style-type: none"> • Reading Comprehension • Writing 	<ul style="list-style-type: none"> • Reading Comprehension • Writing 	<ul style="list-style-type: none"> • Reading Comprehension • Writing 	<ul style="list-style-type: none"> • Reading Comprehension • Writing 	<ul style="list-style-type: none"> • Reading Comprehension • Writing 																	
Mathematics	<ul style="list-style-type: none"> • Number Sense • Operations • Algebra 	<ul style="list-style-type: none"> • Number Sense • Operations • Algebra 	<ul style="list-style-type: none"> • Number Sense • Operations • Algebra 	<ul style="list-style-type: none"> • Number Sense • Operations • Algebra 	<ul style="list-style-type: none"> • Number Sense • Operations • Algebra 																	
Science	<ul style="list-style-type: none"> • Earth Science • Life Science • Physical Science 	<ul style="list-style-type: none"> • Earth Science • Life Science • Physical Science 	<ul style="list-style-type: none"> • Earth Science • Life Science • Physical Science 	<ul style="list-style-type: none"> • Earth Science • Life Science • Physical Science 	<ul style="list-style-type: none"> • Earth Science • Life Science • Physical Science 																	
Social Studies	<ul style="list-style-type: none"> • History • Geography • Civics 	<ul style="list-style-type: none"> • History • Geography • Civics 	<ul style="list-style-type: none"> • History • Geography • Civics 	<ul style="list-style-type: none"> • History • Geography • Civics 	<ul style="list-style-type: none"> • History • Geography • Civics 																	
CONNECTION TO GRADE LEVEL CONTENT = PERFORMANCE = RATING																						
Connections to Grade Level Content = AGI's are the expansion of the academic content for students with severe cognitive disabilities. The assessment tasks must connect to the AGI AND the writing outcome must demonstrate the task. If these connections are not clear, the AGI will not be accepted.																						
Performance = Level of Accuracy = Level of Independence																						
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Rating</th> <th>4</th> <th>3</th> <th>2</th> <th>1</th> <th>No Score (NS)</th> </tr> </thead> <tbody> <tr> <td>Level of Accuracy</td> <td> The student demonstrates a high level of accuracy in all responses. </td> <td> The student demonstrates a moderate level of accuracy in all responses. </td> <td> The student demonstrates a low level of accuracy in all responses. </td> <td> The student demonstrates minimal accuracy in all responses. </td> <td> The student demonstrates no accuracy in all responses. </td> </tr> <tr> <td>Level of Independence</td> <td> The student demonstrates a high level of independence in all responses. </td> <td> The student demonstrates a moderate level of independence in all responses. </td> <td> The student demonstrates a low level of independence in all responses. </td> <td> The student demonstrates minimal independence in all responses. </td> <td> The student demonstrates no independence in all responses. </td> </tr> </tbody> </table>					Rating	4	3	2	1	No Score (NS)	Level of Accuracy	The student demonstrates a high level of accuracy in all responses.	The student demonstrates a moderate level of accuracy in all responses.	The student demonstrates a low level of accuracy in all responses.	The student demonstrates minimal accuracy in all responses.	The student demonstrates no accuracy in all responses.	Level of Independence	The student demonstrates a high level of independence in all responses.	The student demonstrates a moderate level of independence in all responses.	The student demonstrates a low level of independence in all responses.	The student demonstrates minimal independence in all responses.	The student demonstrates no independence in all responses.
Rating	4	3	2	1	No Score (NS)																	
Level of Accuracy	The student demonstrates a high level of accuracy in all responses.	The student demonstrates a moderate level of accuracy in all responses.	The student demonstrates a low level of accuracy in all responses.	The student demonstrates minimal accuracy in all responses.	The student demonstrates no accuracy in all responses.																	
Level of Independence	The student demonstrates a high level of independence in all responses.	The student demonstrates a moderate level of independence in all responses.	The student demonstrates a low level of independence in all responses.	The student demonstrates minimal independence in all responses.	The student demonstrates no independence in all responses.																	

NYSAA Administration Manual 2007-08 (September 2007) 1

Student Performance Data

Level of Accuracy

Number of correct responses

20 correct responses out of 25 possible total responses
 $20/25 = .8 \times 100\% = 80\%$ accurate = Level 4

Level of Independence

Frequency of prompts per number of steps or items in the task

Student completes a 6 step task with prompts on 2 steps $4/6 = .66667$, rounded to $.67 \times 100\% = 67\%$ independent = Level 3

Datafolios during Standard Setting

- Actual student datafolios
- Compiled during the 2007-08 administration
- Scored in the Spring of 2008
- Single content areas only

Slide 1

**New York Student
Alternate Assessment**

Setting Alternate Achievement
Standards
June 9-13, 2008



Slide 2

Purpose of Standard Setting

- Provide data to establish cut scores for:
 - ELA and Math, Grades 3-8 & HS
 - Science, Grades 4, 8 & HS
 - Social Studies, Grades 5, 8 & HS
- Some groups will be recommending cuts for two different grade levels and will be working Monday through Friday:
 - ELA and Math, Grades 3-8
 - Science and Social Studies, Grades 8 & HS
- The remaining groups will be recommending cuts for one grade level and will be working Monday through Wednesday.

2

Slide 3

Purpose of Standard Setting

- Not Meeting the Learning Standards* ← Cut 1
- Partially Meeting the Learning Standards* ← Cut 2
- Meeting the Learning Standards* ← Cut 3
- Meeting the Learning Standards with Distinction*

3

Slide 4

What is Standard Setting?

- Set of activities that result in the determination of threshold or cut scores on an assessment
- We are trying to answer the question:
 - How much is enough?

4

Slide 5

What is Standard Setting?

- Data collection phase
- Policy/Decision making phase

5

Slide 6

Many Standard Setting Methods

- Angoff
- Body of Work
- Bookmark

6

Slide 7

Choice of Method is Based on Many Factors

- Prior usage/history
- Recommendation/requirement by some policy making authority
- Type of assessment

7

Slide 8

Modified Body of Work

- Is especially useful for assessments that consist primarily or entirely of constructed-response items.
- Has been used successfully by Measured Progress in the past.
- Allows panelists to use samples of actual student work to make their determinations.

8

Slide 9

Modified Body of Work Method

- Panelists will be basing their decisions on a set of 20 to 23 student datafolios.
- Datafolios cover the range of possible scores and are presented in order from lowest to highest total score.

9

Slide 10

What is your role in this process?

- To classify each datafolio into the performance level in which you feel it belongs:
 - *Not Meeting the Learning Standards*
 - *Partially Meeting the Learning Standards*
 - *Meeting the Learning Standards*
 - *Meeting the Learning Standards with Distinction*

10

Slide 11

What is your role in this process?

- You may disagree about the order of the datafolios; that's fine.
- You will categorize the datafolios as you see fit, whether your ratings agree with the order or not.
- However, it is not your job to rescore the datafolios: you need to stay focused on the task at hand.

11

Slide 12

Modified Body of Work Method

- Prior to beginning the process of rating the datafolios, you will:
 - thoroughly review and discuss the Alternate Performance Level Descriptors (APLDs);
 - create bulleted lists on chart paper of the knowledge, skills and abilities that a student must demonstrate in order to be categorized into a given performance level.
- It is critical that panelists come to a common understanding of the APLDs.

12

Slide 13

Overview

- Middle Cut Ratings
 - Round 1 (individual)
 - Round 2 (group)
- Lower Cut Ratings
 - Round 1 (individual)
 - Round 2 (group)
- Upper Cut Ratings
 - Round 1 (individual)
 - Round 2 (group)
- Round 3 Ratings (all three cuts)

13

Slide 14

Steps for Modified Body of Work Method

- Round 1
 - Panelists individually review the datafolios.
 - There is no discussion with colleagues.
 - Panelists make their first set of ratings.
- Round 2
 - All panelists in the group will discuss the Round 1 ratings.
 - Panelists make their second set of ratings.

14

Slide 15

Steps for Modified Body of Work Method

- Rounds 1 and 2 will be completed first for the middle cut (below *Meeting the Learning Standards vs. Meeting the Learning Standards* or above).
- Rounds 1 and 2 will next be completed for the lower cut (*Not Meeting the Learning Standards vs. Meeting the Learning Standards*).
- Finally, Rounds 1 and 2 will be completed for the upper cut (*Meeting the Learning Standards vs. Meeting the Learning Standards with Distinction*).

15

Slide 16

Steps for Modified Body of Work Method

- Once Rounds 1 and 2 have been completed for all three cuts, Round 3 occurs:
 - Group discussion of the Round 2 ratings
 - Look at all three cuts simultaneously: more holistic approach
 - You will also be given impact data, indicating the percentage of students who would fall into each category according to the Round 2 ratings.
 - Final round of ratings

16

Slide 17

Steps for Modified Body of Work Method

- For those groups setting standards for two grade levels, you will then repeat the entire process for the second grade:
 - Discuss APLDs & create bulleted lists
 - Round 1-3 Ratings

17

Slide 18

Steps for Modified Body of Work Method

- Note:
 - Your group does not need to come to consensus about how the datafolios should be categorized.
 - You may change your ratings as a result of the discussions, or you may not.
 - You should be open-minded when listening to your colleagues' rationales for their ratings.
 - However: we want your **individual best judgment** in each round of rating.

18

Slide 19

Steps for Modified Body of Work Method

- Note also:
 - This session is intended to be an overview.
 - Your room facilitator will give you lots more details and will guide you through the process step by step.

19

Slide 20

Any Questions about the Modified Body of Work Procedure?

Slide 21

What Next?

- After this session, you will break into grade level/ content area groups

21

Slide 22

What Next?

- Once in your breakout room, you will:
 - review the Alternate Performance Level Descriptors and create your bulleted lists;
 - complete Rounds 1 & 2 for the middle cut;
 - complete Rounds 1 & 2 for the lower cut;
 - complete Rounds 1 & 2 for the upper cut;
 - complete Round 3 for all three cuts.

22

Slide 23

What Next?

- As the final step, we will ask you to complete an evaluation of the standard-setting process.
 - Your honest feedback is important for us, both for improving future standard settings, and for evaluating the results of this one.

23

Slide 24

Thank you for your participation!

APPENDIX D—GROUP FACILITATOR’S SCRIPTS AND OUTLINES

GENERAL INSTRUCTIONS FOR GROUP FACILITATORS NYSAA STANDARD SETTING

ELA & Mathematics, Grades 3-8 Science & Social Studies, Grade 8 and High School

June 9-13, 2008

Introductions

- 1) Welcome group, introduce yourself (name, affiliation, a little selected background information).
- 2) Have each participant introduce him/herself.
- 3) Have participants complete the non-disclosure forms.

Discuss Alternate Performance Level Descriptors (APLDs)

Overview: In order to establish a thorough understanding of the expected performance of students on the test, panelists must have a clear understanding of:

- 1) The definition of the four performance levels, and
- 2) What the key characteristics are that distinguish students in adjacent performance level categories.

The purpose of this activity is for the panelists to come to consensus about what characterizes students in each of the four performance level categories. This activity is critical since the ratings panelists will be making in Rounds 1 through 3 will be based on these understandings.

Activities:

1. Introduce task. In this activity they will:
 - a. Individually review the Alternate Performance Level Descriptors;
 - b. discuss Descriptors as a group;
 - c. Generate bulleted lists that describe the main characteristics that define students in each performance level category.

2. Have panelists individually review all Alternate Performance Level Descriptors (APLD). They can make notes if they like. The goal here is for the panelists to come to a common understanding of what it means to be in each performance level. It is not unusual for panelists to disagree with the descriptions they will see; almost certainly there will be some panelists who will want to change them. However, the task at hand is for panelists to have a common understanding of what knowledge, skills, and understanding are described by each Descriptor. Panelists will be given an opportunity at the end of the process to provide feedback on the Descriptors.
3. After individually reviewing the Descriptors, have the panelists discuss each one as a group, starting with *Partially Meeting the Learning Standards*, and provide clarification. Assign a note taker to capture the points made and any questions. The purpose of this is to have a collegial discussion in which to bring up/clarify any issues or questions that any individual may have and to reach consensus on an understanding of the description.
4. Once the discussion is complete for each performance level, using chart paper, create a bulleted list for each level, specifying the characteristics that best describe students in that level. The panelists want to answer the question, what characteristics must a student demonstrate in order to be classified in the *Partially Meeting the Learning Standards* category. Or, put another way, what are the most important characteristics that distinguish a *Not Meeting the Learning Standards* student from a student in the *Partially Meeting the Learning Standards* category. They will then repeat this process for the *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction* categories.

Ratings: Middle Cut

Overview of Middle Cut Ratings: The panelists will begin the rating process by separating the datafolios into two piles, those that represent performance that is below *Meeting the Learning Standards* (*Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*) vs. *Meeting the Learning Standards* or above (*Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*). The ratings will be done in two rounds. The first round will be done individually, without consulting with their colleagues. In the second round, they will have an opportunity to discuss their Round 1 ratings with the other panelists.

Middle Cut Round 1: The first step in the process will be for the panelists to individually review the datafolios, beginning with #1, and then every fifth datafolio after that (i.e., #6, #11, etc.). Once they have narrowed in on the datafolios they feel are near the cut point between below *Meeting the Learning Standards* and *Meeting the Learning Standards* or above, they will review all the datafolios in that range. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is below *Meeting the Learning Standards*, or *Meeting the Learning Standards* or above. At the end of Round 1, each panelist will complete the Round 1 section of the Middle Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
2. Orient panelists to the set of datafolios. Explain that the datafolios are ordered by the student's total raw score. Make sure they know that, if they disagree with the order of the datafolios, they are free to categorize them as they feel appropriate, regardless of their ordering. For example, if they feel that datafolio #15 represents performance that is *Meeting the Learning Standards* or

above, but #16 (which has a higher total score) represents below *Meeting the Learning Standards* performance, they should categorize them as such.

3. Provide an overview of Round 1. Emphasize the following:
 - a. The primary purpose is to separate the datafolios into two piles.
 - b. Panelists will be working individually in this round, without consulting with their colleagues. They will have opportunities in Rounds 2 and 3 to discuss their categorizations and make changes.
 - c. Each panelist needs to base his/her judgments on his/her experience with the content, understanding of students, and the Descriptors.
 - d. If panelists are struggling with categorizing a particular datafolio, they should use their best judgment and move on. They will have an opportunity to revise their categorizations.
 - e. Panelists should feel free to take notes if there are particular points about a certain datafolio and how they think it should be categorized that they would like to discuss in Round 2.
4. Go over the rating form with panelists:
 - a. Have panelists write their ID number on the rating form. The ID number is on their name tags.
 - b. Lead panelists through a step-by-step demonstration of how to fill in the rating form.
 - c. There should be one and only one checkmark in each row for each round of ratings.
5. Give panelists an opportunity to ask questions about their task in Round 1, then tell them they may begin.
6. Have panelists individually review the datafolios, beginning with #1, and then every fifth one after that (i.e., #6, #11, etc.), ending with the last datafolio. It is important that panelists continue all the way through the last datafolio so they have a good sense of the entire range of performance represented. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: below *Meeting the Learning Standards*, vs. *Meeting the Learning Standards* or above.
7. Once they have narrowed in on the datafolios they feel are near the cut point between below *Meeting the Learning Standards* and *Meeting the Learning Standards* or above, they will review all the datafolios in that range, again placing each in the appropriate pile. **Note:** the panelists will not be reviewing all the datafolios at this time; this is done intentionally, to break the work into more manageable pieces.
8. Panelists may want to take notes as they work.
9. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Middle Cut Rating Form.
10. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one category.
 - c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Middle Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
2. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category (below *Meeting the Learning Standards* vs. *Meeting the Learning Standards* or above).
3. Beginning with the first datafolio for which there was disagreement as to its categorization, the panelists will discuss their rationale for categorizing it as they did.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to how they should be categorized.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

4. Once the discussions have been completed, each panelist will complete the Round 2 section of the Middle Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.
5. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number is filled in.
 - b. Each datafolio must have one (and only one) rating.
 - c. Check each panelist's form before allowing them to leave for a short break.
 - d. When all the forms have been collected, bring them to the R&A work room for tabulation.

Ratings: Lower Cut

Overview of Lower Cut Ratings: Once Rounds 1 and 2 have been completed for the middle cut, the process will be repeated for the lower cut. The panelists will set aside the pile of datafolios that they have classified as *Meeting the Learning Standards* or above, and work only with the datafolios they feel are below *Meeting the Learning Standards*. Working their way through each datafolio in the pile, the panelists will subdivide them into two new piles: *Not Meeting the Learning Standards* and *Partially Meeting the Learning Standards*. As with the middle cut ratings, in the first round of ratings, panelists will work individually and, in the second round, they will have an opportunity to discuss their categorizations before making their second round ratings.

Lower Cut Round 1: The process here will be basically the same as for the middle cut, except that they will be subdividing the datafolios they categorized as below *Meeting the Learning Standards* into two performance levels: *Not Meeting the Learning Standards* and *Partially Meeting the Learning Standards*. They will individually work their way through each of the datafolios they categorized as below *Meeting the Learning Standards*. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is *Not Meeting the Learning Standards*, or *Partially Meeting the Learning Standards*. At the end of Round 1, each panelist will complete the Round 1 section of the Lower Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
 - d. Rating Form for the Lower Cut
 - e. Rating Form for the Upper Cut (they will be preparing it for when they get to the upper cut ratings)
2. Ask the panelists to transfer their ratings in the Round 2: *Meeting the Learning Standards* or Above column of the Middle Cut Rating Form into the *Meeting the Learning Standards* or Above columns of the Lower Cut Rating Form; the ratings should be entered into the *Meeting the Learning Standards* or Above column for both rounds. Once they have done that, have them transfer their Below *Meeting the Learning Standards* ratings onto the Upper Cut Rating Form, again placing them in the Below *Meeting the Learning Standards* columns for both rounds.
3. Have the panelists place the pile of datafolios they categorized as above *Meeting the Learning Standards*, as well as the Upper Cut Rating Form, aside, where they will be out of their way.
4. Have the panelists individually review each datafolio in their below *Meeting the Learning Standards* pile; they will have reviewed some of them while doing their middle cut ratings, but they should revisit those briefly to refresh their memory.
5. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*.
6. **Note:** Because the panelists will be reviewing some datafolios for the first time in this step, it is possible that they may feel that one or more should have been placed in the *Meeting the Learning Standards* or above pile in the previous step. Tell them that, in that case, they should categorize it as *Partially Meeting the Learning Standards* for the time being, but make a note on it indicating that it needs to be recategorized. They will have an opportunity in Round 3 to change any of the categorizations; for now, however, they may not move datafolios out of the below *Meeting the Learning Standards* category.
6. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Lower Cut Rating Form.
7. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one performance level.

- c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Lower Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Lower Cut
2. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category. In this case, you will be including three categories: *Not Meeting the Learning Standards*, *Partially Meeting the Learning Standards*, and *Meeting the Learning Standards* or above. Even though the panelists will be confining their discussions to the *Not Meeting the Learning Standards/Partially Meeting the Learning Standards* cut, including all three categories on the chart paper should help minimize any confusion.
3. Beginning with the first datafolio for which there was disagreement as to whether it should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, the panelists will discuss their rationale for categorizing it as they did.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to whether it should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

4. Once the discussions have been completed, each panelist will complete the Round 2 section of the Lower Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.
5. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number is filled in.
 - b. Each datafolio must have one (and only one) rating.
 - c. Check each panelist's form before allowing them to leave for a short break.

- d. When all the forms have been collected, bring them to the R&A work room for tabulation.

Ratings: Upper Cut

Overview of Upper Cut Ratings: Once Rounds 1 and 2 have been completed for the middle and lower cuts, the process will be repeated one more time for the upper cut. The panelists will set aside the two piles of datafolios that they have classified as either *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, and work only with the datafolios they feel are *Meeting the Learning Standards* or above. Working their way through each datafolio in the pile, the panelists will subdivide them into two new piles: *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction*. As with the middle and lower cut ratings, in the first round of ratings, panelists will work individually and, in the second round, they will have an opportunity to discuss their categorizations before making their second round ratings.

Upper Cut Round 1: The process here will be basically the same as for the lower cut, except that they will be subdividing the datafolios they categorized as *Meeting the Learning Standards* or above into two performance levels: *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction*. They will individually work their way through each of the datafolios they categorized as *Meeting the Learning Standards* or above. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is *Meeting the Learning Standards*, or *Meeting the Learning Standards with Distinction*. At the end of Round 1, each panelist will complete the Round 1 section of the Upper Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Upper Cut
2. Have the panelists place the piles of datafolios they categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards* aside, where they will be out of their way.
3. Have the panelists individually review each datafolio in their *Meeting the Learning Standards* or above pile; they will have reviewed some of them while doing their middle cut ratings, but they should revisit those briefly to refresh their memory.
4. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*.
5. **Note:** Because the panelists will be reviewing some datafolios for the first time in this step, it is possible that they may feel that one or more should have been placed in the below *Meeting the Learning Standards* pile in the first step. Tell them that, in that case, they should categorize it as *Meeting the Learning Standards* for the time being, but make a note on it indicating that it needs to be recategorized. They will have an opportunity in Round 3 to change any of the categorizations; for now, however, they may not move datafolios out of the *Meeting the Learning Standards* or above category.
6. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Upper Cut Rating Form.

7. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one performance level.
 - c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Upper Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Upper Cut
2. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category. In this case, you will be including three categories: below *Meeting the Learning Standards*, *Meeting the Learning Standards*, and *Meeting the Learning Standards with Distinction*. Even though the panelists will be confining their discussions to the *Meeting the Learning Standards/Meeting the Learning Standards with Distinction* cut, including all three categories on the chart paper should help minimize any confusion.
3. Beginning with the first datafolio for which there was disagreement as to whether it should be categorized as *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*, the panelists will discuss their rationale for categorizing it as they did.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to whether they should be categorized as *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

4. Once the discussions have been completed, each panelist will complete the Round 2 section of the Upper Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.

5. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number is filled in.
 - b. Each datafolio must have one (and only one) rating.
 - c. Check each panelist's form before allowing them to leave for a short break.
 - d. When all the forms have been collected, bring them to the R&A work room for tabulation.

Tabulation of Round 2 Results

Once Round 2 has been completed for all three cuts, the data will be analyzed and information will be provided that the panelists will use for Round 3.

Ratings: Round 3 – All Cuts

Overview of Round 3: The primary purpose of Round 3 is to ask the panelists to discuss their Round 2 ratings for all three cuts as a whole group and to revise their ratings on the basis of that discussion. They will discuss their ratings in the context of the ratings made by other members of the group. Prior to beginning the Round 3 discussions, using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each of the four performance level categories. Also show on the chart paper which datafolios will be assigned to each level according to the group average cut points from Round 2 (you will be provided this information by the data analysis team). Focusing on the datafolios that are near the cut points, the panelists will discuss why they categorized each datafolio as they did, making sure that all different points of view are included in the discussion.

To aid with the discussion, panelists will also be given two additional pieces of information. The first will be a graphic display, showing the cuts assigned by each individual panelist. The facilitator will receive the information from the data analysis staff, and will write it up on chart paper, using an alphabetic code for each panelist to maintain anonymity. This information will show panelists whether their cutpoint placements are in line with those of the other panelists or one or more of them is an outlier. The second piece of information will be impact data, showing the approximate percentage of students who would be classified into each performance level category based on the room average cutpoints from Round 2.

This round will be similar to the Round 2 discussions, except that the panelists will be discussing all three cut points. The purpose of this round is to look at the results holistically, rather than each cut individually. Therefore, the panelists should start the discussions with the lower cut, then proceed to the middle cut and, finally, the upper cut.

Once panelists have reviewed and discussed the Round 2 categorizations, they will be given the opportunity to change or revise their Round 2 ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. The Round 3 rating form
 - b. Set of datafolios

- c. Alternate Performance Level Descriptors
2. Have panelists write their ID number on the rating form.
3. Provide an overview of Round 3. Paraphrase the following:
 - a. As in Rounds 1 and 2, the primary purpose is to categorize each datafolio into the alternate performance level category where you feel it belongs.
 - b. Each panelist needs to base his/her judgments on his/her experience with the content area, understanding of students, discussions with other panelists and the knowledge, skills,, and understanding required to answer each item.
 - c. In addition to the categorization of each datafolio, panelists should also consider the impact data: based on their knowledge of students and the Descriptors, do the percentages of students falling into each category make sense? If they do, that is an indication that the cutpoints are placed appropriately. If they don't, the panelists may want to consider revising their ratings.
4. Review the feedback information with the panelists.
 - a. Show the panelists how the datafolios will be categorized based on the room average Round 2 cut point placements.
 - b. Go over the chart of each panelist's cut-point placements. Tell panelists that if one or more of their cuts is substantially higher or lower than those of their colleagues, they should revisit their understanding of the Descriptors and their rationale for their placement(s).
 - c. Go over the impact data, explaining that if the Round 2 ratings were to be used to set the final cutpoints, these are the approximate percentages of students who would be classified into each alternate performance level category.
5. Give panelists an opportunity to ask questions about the feedback information or about the task for Round 3.
6. Beginning with the datafolios for which there was disagreement as to whether they should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, the panelists should begin discussing the categorization of the datafolios according to the Round 2 ratings. Once they have completed the discussion for the lower cut, they will then proceed to the middle cut and then, finally, to the upper cut.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to how they should be categorized.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should place it into one of four piles: *Not Meeting the Learning Standards*, *Partially Meeting the Learning Standards*, *Meeting the Learning Standards*, or *Meeting the Learning Standards with Distinction*.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

7. Once the discussions are complete for the full set of datafolios, have the panelists fill in the Round 3 Rating Form. When you collect the rating forms, carefully inspect them to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio for Round 3 must have one (and only one) rating.

Recommendations for enhancements or modifications to Alternate Performance Level Descriptors

Upon completion of Round 3, if time allows, the panelists will have an opportunity to provide feedback on the Descriptors. These recommendations are usually centered on providing language that is clearer and more teacher- and parent-friendly. Make sure panelists know that these are recommendations and that they may not all be implemented.

Complete Entire Standard-setting Process for Second Grade Level

Once the group has completed all standard-setting activities for the first grade level, they will repeat the entire process for the second grade level, beginning with reviewing the APLDs and creating the bulleted lists, and ending with the recommendations for the APLDs for the second grade level.

Complete Evaluation Form

Upon completion of the standard-setting activities for both grade levels, have panelists fill out the evaluation form. Emphasize that their honest feedback is important.

GENERAL INSTRUCTIONS FOR GROUP FACILITATORS NYSAA STANDARD SETTING

ELA & Mathematics, High School Science, Grade 4 Social Studies, Grade 5

June 9-11, 2008

Introductions

- 4) Welcome group, introduce yourself (name, affiliation, a little selected background information).
- 5) Have each participant introduce him/herself.
- 6) Have participants complete the non-disclosure forms.

Discuss Alternate Performance Level Descriptors (APLDs)

Overview: In order to establish a thorough understanding of the expected performance of students on the test, panelists must have a clear understanding of:

- 3) The definition of the four performance levels, and
- 4) what the key characteristics are that distinguish students in adjacent performance level categories.

The purpose of this activity is for the panelists to come to consensus about what characterizes students in each of the four performance level categories. This activity is critical since the ratings panelists will be making in Rounds 1 through 3 will be based on these understandings.

Activities:

5. Introduce task. In this activity they will:
 - d. Individually review the Alternate Performance Level Descriptors;
 - e. discuss Descriptors as a group;
 - f. generate bulleted lists that describe the main characteristics that define students in each performance level category.

6. Have panelists individually review all Alternate Performance Level Descriptors (APLD). They can make notes if they like. The goal here is for the panelists to come to a common understanding of what it means to be in each performance level. It is not unusual for panelists to disagree with the descriptions they will see; almost certainly there will be some panelists who will want to change them. However, the task at hand is for panelists to have a common understanding of what knowledge, skills, and understanding are described by each Descriptor. Panelists will be given an opportunity at the end of the process to provide feedback on the Descriptors.
7. After individually reviewing the Descriptors, have the panelists discuss each one as a group, starting with *Partially Meeting the Learning Standards*, and provide clarification. Assign a note taker to capture the points made and any questions. The purpose of this is to have a collegial discussion in which to bring up/clarify any issues or questions that any individual may have and to reach consensus on an understanding of the description.
8. Once the discussion is complete for each performance level, using chart paper, create a bulleted list for each level, specifying the characteristics that best describe students in that level. The panelists want to answer the question, what characteristics must a student demonstrate in order to be classified in the *Partially Meeting the Learning Standards* category. Or, put another way, what are the most important characteristics that distinguish a *Not Meeting the Learning Standards* student from a student in the *Partially Meeting the Learning Standards* category. They will then repeat this process for the *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction* categories.

Ratings: Middle Cut

Overview of Middle Cut Ratings: The panelists will begin the rating process by separating the datafolios into two piles, those that represent performance that is below *Meeting the Learning Standards* (*Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*) vs. *Meeting the Learning Standards* or above (*Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*). The ratings will be done in two rounds. The first round will be done individually, without consulting with their colleagues. In the second round, they will have an opportunity to discuss their Round 1 ratings with the other panelists.

Middle Cut Round 1: The first step in the process will be for the panelists to individually review the datafolios, beginning with #1, and then every fifth datafolio after that (i.e., #6, #11, etc.). Once they have narrowed in on the datafolios they feel are near the cut point between below *Meeting the Learning Standards* and *Meeting the Learning Standards* or above, they will review all the datafolios in that range. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is below *Meeting the Learning Standards*, or *Meeting the Learning Standards* or above. At the end of Round 1, each panelist will complete the Round 1 section of the Middle Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

11. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
12. Orient panelists to the set of datafolios. Explain that the datafolios are ordered by the student's total raw score. Make sure they know that, if they disagree with the order of the datafolios, they are free to categorize them as they feel appropriate, regardless of their ordering. For example, if they feel that datafolio #15 represents performance that is *Meeting the Learning Standards* or

above, but #16 (which has a higher total score) represents below *Meeting the Learning Standards* performance, they should categorize them as such.

13. Provide an overview of Round 1. Emphasize the following:
 - a. The primary purpose is to separate the datafolios into two piles.
 - b. Panelists will be working individually in this round, without consulting with their colleagues. They will have opportunities in Rounds 2 and 3 to discuss their categorizations and make changes.
 - c. Each panelist needs to base his/her judgments on his/her experience with the content, understanding of students, and the Descriptors.
 - d. If panelists are struggling with categorizing a particular datafolio, they should use their best judgment and move on. They will have an opportunity to revise their categorizations.
 - e. Panelists should feel free to take notes if there are particular points about a certain datafolio and how they think it should be categorized that they would like to discuss in Round 2.
14. Go over the rating form with panelists:
 - a. Have panelists write their ID number on the rating form. The ID number is on their name tags.
 - b. Lead panelists through a step-by-step demonstration of how to fill in the rating form.
 - c. There should be one and only one checkmark in each row for each round of ratings.
15. Give panelists an opportunity to ask questions about their task in Round 1, then tell them they may begin.
16. Have panelists individually review the datafolios, beginning with #1, and then every fifth one after that (i.e., #6, #11, etc.), ending with the last datafolio. It is important that panelists continue all the way through the last datafolio so they have a good sense of the entire range of performance represented. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: below *Meeting the Learning Standards*, vs. *Meeting the Learning Standards* or above.
17. Once they have narrowed in on the datafolios they feel are near the cut point between below *Meeting the Learning Standards* and *Meeting the Learning Standards* or above, they will review all the datafolios in that range, again placing each in the appropriate pile. **Note:** the panelists will not be reviewing all the datafolios at this time; this is done intentionally, to break the work into more manageable pieces.
18. Panelists may want to take notes as they work.
19. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Middle Cut Rating Form.
20. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one category.
 - c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Middle Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

8. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
9. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category (below *Meeting the Learning Standards* vs. *Meeting the Learning Standards* or above).
10. Beginning with the first datafolio for which there was disagreement as to its categorization, the panelists will discuss their rationale for categorizing it as they did.
 - f. Panelists only need to discuss those datafolios for which there was disagreement as to how they should be categorized.
 - g. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - h. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - i. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - j. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

11. Once the discussions have been completed, each panelist will complete the Round 2 section of the Middle Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.
12. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number is filled in.
 - b. Each datafolio must have one (and only one) rating.
 - c. Check each panelist's form before allowing them to leave for a short break.
 - d. When all the forms have been collected, bring them to the R&A work room for tabulation.

Ratings: Lower Cut

Overview of Lower Cut Ratings: Once Rounds 1 and 2 have been completed for the middle cut, the process will be repeated for the lower cut. The panelists will set aside the pile of datafolios that they have classified as *Meeting the Learning Standards* or above, and work only with the datafolios they feel are below *Meeting the Learning Standards*. Working their way through each datafolio in the pile, the panelists will subdivide them into two new piles: *Not Meeting the Learning Standards* and *Partially Meeting the Learning Standards*. As with the middle cut ratings, in the first round of ratings, panelists will work individually and, in the second round, they will have an opportunity to discuss their categorizations before making their second round ratings.

Lower Cut Round 1: The process here will be basically the same as for the middle cut, except that they will be subdividing the datafolios they categorized as below *Meeting the Learning Standards* into two performance levels: *Not Meeting the Learning Standards* and *Partially Meeting the Learning Standards*. They will individually work their way through each of the datafolios they categorized as below *Meeting the Learning Standards*. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is *Not Meeting the Learning Standards*, or *Partially Meeting the Learning Standards*. At the end of Round 1, each panelist will complete the Round 1 section of the Lower Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

7. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Middle Cut
 - d. Rating Form for the Lower Cut
 - e. Rating Form for the Upper Cut (they will be preparing it for when they get to the upper cut ratings)
8. Ask the panelists to transfer their ratings in the Round 2: *Meeting the Learning Standards* or Above column of the Middle Cut Rating Form into the *Meeting the Learning Standards* or Above columns of the Lower Cut Rating Form; the ratings should be entered into the *Meeting the Learning Standards* or Above column for both rounds. Once they have done that, have them transfer their Below *Meeting the Learning Standards* ratings onto the Upper Cut Rating Form, again placing them in the Below *Meeting the Learning Standards* columns for both rounds.
9. Have the panelists place the pile of datafolios they categorized as above *Meeting the Learning Standards*, as well as the Upper Cut Rating Form, aside, where they will be out of their way.
10. Have the panelists individually review each datafolio in their below *Meeting the Learning Standards* pile; they will have reviewed some of them while doing their middle cut ratings, but they should revisit those briefly to refresh their memory.
11. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*.
12. **Note:** Because the panelists will be reviewing some datafolios for the first time in this step, it is possible that they may feel that one or more should have been placed in the *Meeting the Learning Standards* or above pile in the previous step. Tell them that, in that case, they should categorize it as *Partially Meeting the Learning Standards* for the time being, but make a note on it indicating that it needs to be recategorized. They will have an opportunity in Round 3 to change any of the categorizations; for now, however, they may not move datafolios out of the below *Meeting the Learning Standards* category.
13. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Lower Cut Rating Form.
14. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one performance level.

- c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Lower Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

6. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Lower Cut
7. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category. In this case, you will be including three categories: *Not Meeting the Learning Standards*, *Partially Meeting the Learning Standards*, and *Meeting the Learning Standards* or above. Even though the panelists will be confining their discussions to the *Not Meeting the Learning Standards/Partially Meeting the Learning Standards* cut, including all three categories on the chart paper should help minimize any confusion.
8. Beginning with the first datafolio for which there was disagreement as to whether it should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, the panelists will discuss their rationale for categorizing it as they did.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to whether it should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

9. Once the discussions have been completed, each panelist will complete the Round 2 section of the Lower Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.
10. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - e. The ID number is filled in.
 - f. Each datafolio must have one (and only one) rating.
 - g. Check each panelist's form before allowing them to leave for a short break.

- h. When all the forms have been collected, bring them to the R&A work room for tabulation.

Ratings: Upper Cut

Overview of Upper Cut Ratings: Once Rounds 1 and 2 have been completed for the middle and lower cuts, the process will be repeated one more time for the upper cut. The panelists will set aside the two piles of datafolios that they have classified as either *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, and work only with the datafolios they feel are *Meeting the Learning Standards* or above. Working their way through each datafolio in the pile, the panelists will subdivide them into two new piles: *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction*. As with the middle and lower cut ratings, in the first round of ratings, panelists will work individually and, in the second round, they will have an opportunity to discuss their categorizations before making their second round ratings.

Upper Cut Round 1: The process here will be basically the same as for the lower cut, except that they will be subdividing the datafolios they categorized as *Meeting the Learning Standards* or above into two performance levels: *Meeting the Learning Standards* and *Meeting the Learning Standards with Distinction*. They will individually work their way through each of the datafolios they categorized as *Meeting the Learning Standards* or above. As they proceed through the datafolios, the panelists should ask themselves whether the knowledge, skills, and understanding demonstrated in each are consistent with performance that is *Meeting the Learning Standards*, or *Meeting the Learning Standards with Distinction*. At the end of Round 1, each panelist will complete the Round 1 section of the Upper Cut Rating Form, indicating the level they feel each datafolio should be categorized into.

Activities:

8. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Upper Cut
9. Have the panelists place the piles of datafolios they categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards* aside, where they will be out of their way.
10. Have the panelists individually review each datafolio in their *Meeting the Learning Standards* or above pile; they will have reviewed some of them while doing their middle cut ratings, but they should revisit those briefly to refresh their memory.
11. As they are reviewing the datafolios, the panelists should keep in mind the Descriptors. They should consider the knowledge, skills, and understanding demonstrated by each and how they relate to the definitions of the performance levels. As they complete each datafolio, have them place it into one of two piles: *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*.
12. **Note:** Because the panelists will be reviewing some datafolios for the first time in this step, it is possible that they may feel that one or more should have been placed in the below *Meeting the Learning Standards* pile in the first step. Tell them that, in that case, they should categorize it as *Meeting the Learning Standards* for the time being, but make a note on it indicating that it needs to be recategorized. They will have an opportunity in Round 3 to change any of the categorizations; for now, however, they may not move datafolios out of the *Meeting the Learning Standards* or above category.
13. Once panelists have finished sorting the datafolios, they will fill in the Round 1 section of the Upper Cut Rating Form.

14. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio must be assigned to one and only one performance level.
 - c. Although the datafolios are presented in order from lowest- to highest-scoring, the panelists' category assignments do not need to be in strictly increasing order.

Upper Cut Round 2: In Round 2, the panelists will discuss their categorizations of the datafolios into the two levels as a large group. After the discussions are complete, the panelists will do their second round of ratings.

Activities:

6. Make sure panelists have the following materials:
 - a. Set of datafolios
 - b. Alternate Performance Level Descriptors
 - c. Rating Form for the Upper Cut
7. Using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each category. In this case, you will be including three categories: below *Meeting the Learning Standards*, *Meeting the Learning Standards*, and *Meeting the Learning Standards with Distinction*. Even though the panelists will be confining their discussions to the *Meeting the Learning Standards/Meeting the Learning Standards with Distinction* cut, including all three categories on the chart paper should help minimize any confusion.
8. Beginning with the first datafolio for which there was disagreement as to whether it should be categorized as *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*, the panelists will discuss their rationale for categorizing it as they did.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to whether they should be categorized as *Meeting the Learning Standards* or *Meeting the Learning Standards with Distinction*.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should once again place it into the appropriate pile.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors than the rest of the group. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

9. Once the discussions have been completed, each panelist will complete the Round 2 section of the Upper Cut Rating Form, again indicating the level they feel each datafolio should be categorized into.

10. As panelists complete the task, ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - e. The ID number is filled in.
 - f. Each datafolio must have one (and only one) rating.
 - g. Check each panelist's form before allowing them to leave for a short break.
 - h. When all the forms have been collected, bring them to the R&A work room for tabulation.

Tabulation of Round 2 Results

Once Round 2 has been completed for all three cuts, the data will be analyzed and information will be provided that the panelists will use for Round 3.

Ratings: Round 3 – All Cuts

Overview of Round 3: The primary purpose of Round 3 is to ask the panelists to discuss their Round 2 ratings for all three cuts as a whole group and to revise their ratings on the basis of that discussion. They will discuss their ratings in the context of the ratings made by other members of the group. Prior to beginning the Round 3 discussions, using a show of hands, indicate on a piece of chart paper how many panelists assigned each datafolio to each of the four performance level categories. Also show on the chart paper which datafolios will be assigned to each level according to the group average cut points from Round 2 (you will be provided this information by the data analysis team). Focusing on the datafolios that are near the cut points, the panelists will discuss why they categorized each datafolio as they did, making sure that all different points of view are included in the discussion.

To aid with the discussion, panelists will also be given two additional pieces of information. The first will be a graphic display, showing the cuts assigned by each individual panelist. The facilitator will receive the information from the data analysis staff, and will write it up on chart paper, using an alphabetic code for each panelist to maintain anonymity. This information will show panelists whether their cutpoint placements are in line with those of the other panelists or one or more of them is an outlier. The second piece of information will be impact data, showing the approximate percentage of students who would be classified into each performance level category based on the room average cutpoints from Round 2.

This round will be similar to the Round 2 discussions, except that the panelists will be discussing all three cut points. The purpose of this round is to look at the results holistically, rather than each cut individually. Therefore, the panelists should start the discussions with the lower cut, then proceed to the middle cut and, finally, the upper cut.

Once panelists have reviewed and discussed the Round 2 categorizations, they will be given the opportunity to change or revise their Round 2 ratings.

Activities:

8. Make sure panelists have the following materials:
 - a. The Round 3 rating form
 - b. Set of datafolios

- c. Alternate Performance Level Descriptors
- 9. Have panelists write their ID number on the rating form.
- 10. Provide an overview of Round 3. Paraphrase the following:
 - a. As in Rounds 1 and 2, the primary purpose is to categorize each datafolio into the alternate performance level category where you feel it belongs.
 - b. Each panelist needs to base his/her judgments on his/her experience with the content area, understanding of students, discussions with other panelists and the knowledge, skills,, and understanding required to answer each item.
 - c. In addition to the categorization of each datafolio, panelists should also consider the impact data: based on their knowledge of students and the Descriptors, do the percentages of students falling into each category make sense? If they do, that is an indication that the cutpoints are placed appropriately. If they don't, the panelists may want to consider revising their ratings.
- 11. Review the feedback information with the panelists.
 - a. Show the panelists how the datafolios will be categorized based on the room average Round 2 cut point placements.
 - b. Go over the chart of each panelist's cut-point placements. Tell panelists that if one or more of their cuts is substantially higher or lower than those of their colleagues, they should revisit their understanding of the Descriptors and their rationale for their placement(s).
 - c. Go over the impact data, explaining that if the Round 2 ratings were to be used to set the final cutpoints, these are the approximate percentages of students who would be classified into each alternate performance level category.
- 12. Give panelists an opportunity to ask questions about the feedback information or about the task for Round 3.
- 13. Beginning with the datafolios for which there was disagreement as to whether they should be categorized as *Not Meeting the Learning Standards* or *Partially Meeting the Learning Standards*, the panelists should begin discussing the categorization of the datafolios according to the Round 2 ratings. Once they have completed the discussion for the lower cut, they will then proceed to the middle cut and then, finally, to the upper cut.
 - a. Panelists only need to discuss those datafolios for which there was disagreement as to how they should be categorized.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.
 - e. As they finish the discussion for each datafolio, each panelist should place it into one of four piles: *Not Meeting the Learning Standards*, *Partially Meeting the Learning Standards*, *Meeting the Learning Standards*, or *Meeting the Learning Standards with Distinction*.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is categorizing datafolios consistently higher or lower than the group, he/she may have a different understanding of the Descriptors. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Descriptors.**

14. Once the discussions are complete for the full set of datafolios, have the panelists fill in the Round 3 Rating Form. When you collect the rating forms, carefully inspect them to ensure they are filled out properly.
 - a. The ID number must be filled in.
 - b. Each datafolio for Round 3 must have one (and only one) rating.

Recommendations for enhancements or modifications to Alternate Performance Level Descriptors

Upon completion of Round 3, if time allows, the panelists will have an opportunity to provide feedback on the Descriptors. These recommendations are usually centered on providing language that is clearer and more teacher- and parent-friendly. Make sure panelists know that these are recommendations and that they may not all be implemented.

Complete Evaluation Form

Upon completion of Round 3, have panelists fill out the evaluation form. Emphasize that their honest feedback is important.

Outline of Standard Setting Process for Facilitators

English Language Arts and Mathematics, Grades 3-8
Science and Social Studies, Grade 8 & High School

First Grade Level (ELA & Mathematics: 3/5/7; Science & Social Studies: 8)

Introductions

Complete non-disclosure form

Discuss Alternate Performance Level Descriptors (first grade)

Create bulleted lists on chart paper

Rounds 1 & 2 Ratings: Middle Cut

Complete Round 1 ratings for Middle Cut

Fill in Round 1 section of Middle Cut rating sheets

Post Round 1 ratings for Middle Cut on chart paper

Group discussion of Round 1 ratings for Middle Cut

Fill in Round 2 section of Middle Cut rating sheets and submit to War Room

Timing: you should at least be ready to begin the Round 2 discussions for the Middle Cut by the end of the day Monday.

Rounds 1 & 2 Ratings: Lower Cut

Complete Round 1 ratings for Lower Cut

Fill in Round 1 section of Lower Cut rating sheets

Post Round 1 ratings for Lower Cut on chart paper

Group discussion of Round 1 ratings for Lower Cut

Fill in Round 2 section of Lower Cut rating sheets and submit to War Room

Rounds 1 & 2 Ratings: Upper Cut

Complete Round 1 ratings for Upper Cut

Fill in Round 1 section of Upper Cut rating sheets

Post Round 1 ratings for Upper Cut on chart paper

Group discussion of Round 1 ratings for Upper Cut

Fill in Round 2 section of Upper Cut rating sheets and submit to War Room

Timing: you should at least be ready to begin the Round 2 discussions for the Upper Cut by the end of the day Tuesday.

Round 3 Ratings

Post Round 2 ratings for all three cuts on chart paper

Group discussion of Round 2 ratings

Fill in Round 3 rating sheets and submit to War Room

Recommendations for enhancements or modifications to Alternate Performance Level Descriptors

Timing: you should be finished with the first grade level by early or mid-afternoon on Wednesday

Second Grade Level (ELA & Mathematics: 4/6/8; Science & Social Studies: High School)

Discuss Alternate Performance Level Descriptors (second grade)

Create bulleted lists on chart paper

Rounds 1 & 2 Ratings: Middle Cut

Complete Round 1 ratings for Middle Cut

Fill in Round 1 section of Middle Cut rating sheets

Post Round 1 ratings for Middle Cut on chart paper

Group discussion of Round 1 ratings for Middle Cut

Fill in Round 2 section of Middle Cut rating sheets and submit to War Room

Rounds 1 & 2 Ratings: Lower Cut

Complete Round 1 ratings for Lower Cut

Fill in Round 1 section of Lower Cut rating sheets

Post Round 1 ratings for Lower Cut on chart paper

Group discussion of Round 1 ratings for Lower Cut

Fill in Round 2 section of Lower Cut rating sheets and submit to War Room

Rounds 1 & 2 Ratings: Upper Cut

Complete Round 1 ratings for Upper Cut

Fill in Round 1 section of Upper Cut rating sheets

Post Round 1 ratings for Upper Cut on chart paper

Group discussion of Round 1 ratings for Upper Cut

Fill in Round 2 section of Upper Cut rating sheets and submit to War Room

Timing: you should have completed Rounds 1 and 2 for all three cuts by the end of the day Thursday.

Round 3 Ratings

Post Round 2 ratings for all three cuts on chart paper

Group discussion of Round 2 ratings

Fill in Round 3 rating sheets and submit to War Room

Recommendations for enhancements or modifications to Alternate Performance Level Descriptors

Timing: you must be finished with all standard-setting activities by noon on Friday

Complete Evaluation Form

Be sure to collect all materials from the panelists before they leave

Outline of Standard Setting Process for Facilitators

English Language Arts and Mathematics, High School
Science, Grade 4
Social Studies, Grade 5

Introductions

Complete non-disclosure form

Discuss Alternate Performance Level Descriptors

Create bulleted lists on chart paper

Rounds 1 & 2 Ratings: Middle Cut

Complete Round 1 ratings for Middle Cut

Fill in Round 1 section of Middle Cut rating sheets

Post Round 1 ratings for Middle Cut on chart paper

Group discussion of Round 1 ratings for Middle Cut

Fill in Round 2 section of Middle Cut rating sheets and submit to War Room

Timing: you should at least be ready to begin the Round 2 discussions for the Middle Cut by the end of the day Monday.

Rounds 1 & 2 Ratings: Lower Cut

Complete Round 1 ratings for Lower Cut

Fill in Round 1 section of Lower Cut rating sheets

Post Round 1 ratings for Lower Cut on chart paper

Group discussion of Round 1 ratings for Lower Cut

Fill in Round 2 section of Lower Cut rating sheets and submit to War Room

Rounds 1 & 2 Ratings: Upper Cut

Complete Round 1 ratings for Upper Cut

Fill in Round 1 section of Upper Cut rating sheets

Post Round 1 ratings for Upper Cut on chart paper

Group discussion of Round 1 ratings for Upper Cut

Fill in Round 2 section of Upper Cut rating sheets and submit to War Room

Timing: you should submit your final set of Round 2 ratings by the end of the day on Tuesday.

Round 3 Ratings

Post Round 2 ratings for all three cuts on chart paper

Group discussion of Round 2 ratings

Fill in Round 3 rating sheets and submit to War Room

Recommendations for enhancements or modifications to Alternate Performance Level Descriptors

Timing: you must be finished with all standard-setting activities by noon on Wednesday

Complete Evaluation Form

Be sure to collect all materials from the panelists before they leave

APPENDIX E—LIST OF PARTICIPANTS

2008 Standard Setting Recruitment

Grade(s)	English Language Arts	Mathematics	Science	Social Science
3	9 panelists	8 panelists		
4			9 panelists	
5	8 panelists	9 panelists		7 panelists
6				
7	9 panelists	8 panelists		
8				
High School	7 panelists	8 panelists	7 panelists	9 panelists

ELA	
<p><u>Grades 3/4</u></p> <ol style="list-style-type: none"> Addieg, Kathryn Barge, Marilyn Bell, Niquae Cole, Mary Lewek, Kathy Murphy, Jane Peterson, Ashley Saramak, Lauren Scully, Jennifer 	<p><u>Grades 5/6</u></p> <ol style="list-style-type: none"> Cutler, Christine Grillo-Mancuso, Zoraida Kness, Christopher Kuzan, Ed MacMasters, Charelle Sanford, Jennifer Schulsberg, Barbara Skolnick, Eleanor
<p><u>Grades 7/8</u></p> <ol style="list-style-type: none"> Arnold, Michelle Beaudion, Susan Bloom, Mary Figuroa, Ruth Heavern, Mary McCaffrey, Bob Soraghan, Christine Turner, Michelle Wadsworth-Miller, Melissa 	<p><u>Grade HS</u></p> <ol style="list-style-type: none"> Becker, Anita Gross, Marion Kowell, Tom MacNeil, Beth Merchant, Peter Riley, Michelle Young, Mickey

MATHEMATICS**Grades 3/4**

1. Allaire, Barbara
2. Dudes, Patricia
3. Flury-Kashmanian, Caroline
4. Glowacki, Susan
5. Johnson, Cindy
6. McDaid, Kathleen
7. Smith, Louise
8. Swzed, Beryl

Grades 5/6

1. Brown, Timothy
2. Cole, John
3. Grob, Cindy
4. Hoeglmeier, Kathy
5. Jarvis, Andrea
6. Kelly, Ann
7. O'Donnell, Patrice
8. Rosowski, Arlene
9. Torre, Linda

Grades 7/8

1. Dallara, Ralph
2. Jackson, Rebecca
3. McPartland, Kim
4. Morris, Mary
5. Pearles, Linda
6. Purdy, Breda
7. Skolick, Warren
8. Whitham, Joshua

Grade HS

1. Hazel, Jeffrey
2. Keaveny, Brandan
3. Kirsch, Steve
4. Laub, Agnes
5. Merchant, Barbara
6. Sahrle, Heather
7. Sulner, Ilisa
8. Wood, Terri

SCIENCE**Grade 4**

1. Boguslawski, Marcia
2. Erlichman, Lauren
3. Garab, Dan
4. Johnson, Christine
5. Mousaw, John
6. Owen, Trudy
7. Rothstein, Iris
8. Volkmuth, Kerri
9. Wakely, Kelly

Grade 8 & HS

1. Bao, James
2. Byrne, Melanie
3. Hand, Jennifer
4. Lesinski, Cherie
5. Lippert, Rhonda
6. Loyche, Danielle
7. Phillips-Russ, Catherine

SOCIAL STUDIES**Grade 5**

1. Cotton, Denise
2. Crawford, Jeanne
3. deSpoelberch, Erin
4. Hayman, Janet
5. Heidt, Kristi
6. Sass, Emily
7. Ward, Stacy

Grade 8 & HS

1. Acevedo, Nelson
2. Fletcher-Smith, Victoria
3. Gibbons, Patrick
4. Gramza, Maryann
5. Jenkins, Carl
6. Longo, Elisa
7. Mannino, Matteo
8. Swan, Pam
9. Walsh, Beth

APPENDIX F—SAMPLE RATING FORMS

NYSAA ELA Grade 3 Rating Form – Middle Cut

	Round 1 (individual)		Round 2 (after discussion)	
	<i>Below Meeting</i>	<i>Meeting or Above</i>	<i>Below Meeting</i>	<i>Meeting or Above</i>
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				

Transcribe these figures into the appropriate columns on the Lower and Upper Cut Rating Forms



Below *Meeting* includes:

NM: *Not Meeting the Learning Standards*

PM: *Partially Meeting the Learning Standards*

Meeting or Above includes:

M: *Meeting the Learning Standards*

MD: *Meeting the Learning Standards with Distinction*

NYSAA ELA Grade 3 Rating Form – Lower Cut

	Round 1 (individual)			Round 2 (after discussion)		
	NM	PM	<i>Meeting or Above</i>	NM	PM	<i>Meeting or Above</i>
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						

Transcribe your Round 2 “*Meeting or Above*” Ratings from the Middle Cut Rating Form Here



Below *Meeting* includes:

NM: *Not Meeting the Learning Standards*

PM: *Partially Meeting the Learning Standards*

Meeting or Above includes:

M: *Meeting the Learning Standards*

MD: *Meeting the Learning Standards with Distinction*

NYSAA ELA Grade 3 Rating Form – Upper Cut

	Round 1 (individual)			Round 2 (after discussion)		
	<i>Below Meeting</i>	M	MD	<i>Below Meeting</i>	M	MD
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						


Transcribe your Round 2 “Below Meeting” ratings from the Middle Cut Rating Form here

Below Meeting includes:

NM: *Not Meeting the Learning Standards*

PM: *Partially Meeting the Learning Standards*

Meeting or Above includes:

M: *Meeting the Learning Standards*

MD: *Meeting the Learning Standards with Distinction*

**NYSAA ELA Grade 3
Rating Form – All Cuts**

	Round 3			
	NM	PM	M	MD
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				

NM: *Not Meeting the Learning Standards*
 PM: *Partially Meeting the Learning Standards*
 M: *Meeting the Learning Standards*
 MD: *Meeting the Learning Standards with Distinction*

APPENDIX G—SAMPLE EVALUATION FORM

Standard Setting Panel

Evaluation Form

Evaluation of the Standard Setting Procedures for the New York State Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the New York State Alternate Assessment? (*Circle one*)
 - A. Very Good
 - B. Good
 - C. Neutral
 - D. Poor
 - E. Very Poor

2. How clear were the Alternate Performance Level Descriptors? (*Circle one*)
 - A. Very Clear
 - B. Clear
 - C. Somewhat Clear
 - D. Not Clear

3. How would you judge the length of time of this meeting for setting alternate performance standards? (*Circle one*)
 - A. About right
 - B. Too little time
 - C. Too much time

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?
 - A. Yes
 - B. No

5. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential)
 - A. The Alternate Performance Level Descriptors

Not at all Influential		Moderately Influential		Very Influential
1	2	3	4	5

 - B. The student datafolios

Not at all Influential		Moderately Influential		Very Influential
1	2	3	4	5

C. Other panelists

Not at all Influential		Moderately Influential		Very Influential
1	2	3	4	5

D. My experience in the field

Not at all Influential		Moderately Influential		Very Influential
1	2	3	4	5

E. Other (please specify _____)

Not at all Influential		Moderately Influential		Very Influential
1	2	3	4	5

6. How could the standard setting process have been improved?

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at all Useful					Very Useful
1	2	3	4	5	

8. Providing additional details to the Alternate Performance Level Descriptors was:

Not at all Useful					Very Useful
1	2	3	4	5	

9. The datafolio rating task was:

Not at all Clear					Very Clear
1	2	3	4	5	

10. The discussion with other panelists was:

Not at all Useful					Very Useful
1	2	3	4	5	

11. The impact data provided was:

Not at all Useful					Very Useful
1	2	3	4	5	

12. I was confident in classifying the student datafolios:

Strongly Disagree					Strongly Agree
1	2	3	4	5	

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process. Use extra paper if necessary.

APPENDIX H—RESULTS OF EVALUATION BY PANELISTS

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	1	2	0	0	5

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• Parent of a severely disabled child	0	0	0	1	0
• Facilitator – as far as keeping us focused on task and stimulating discussion	0	0	0	1	0

6. How could the standard setting process have been improved?

- Our group amended the language of the descriptor for one grade level and not the other which I think caused some confusion.
- Length of unproductive time waiting for the psychometric process to be completed on day 2. It was not until 3:30 that we got our information and we stopped at lunch time.
- Performance Level Descriptors should be reviewed but not changed during this process. Performance Level Descriptors need a separate workshop.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	1	0	2	4

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	0	4	4

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	4	4

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	1	7

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	1	1	6

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	2	6

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- It would be helpful to have a glossary of all the acronyms. Alicia is wonderful. I really appreciated her patience and support.
- I felt that final cut scores for grade 3 seemed high, 53% at MD. It is difficult to truly evaluate the scores since some tasks are less complex than others and I would believe that somehow the tasks should be weighted in order to give a more accurate score. It is also difficult to determine the actual amt. of prompting and is it beyond the scope of the IEP? Lastly, when a student receives a score of 2 or below what are the consequences? Gen. Ed. students will receive AIS what more support, accommodations etc. can be given to the Special Ed. students?
- Alicia is an excellent facilitator. Well organized overall.
- Alicia facilitated our group and did an excellent job explaining and supporting all members of the group. She has a tremendous gift of combining humor and expertise to produce quality work/workers. This experience has provided me with additional information. I enjoyed the learning experience.
- Alicia was an excellent facilitator. We had thoughtful and respectful discussions about our topics and had a great time in the process. Of continual concern when setting the standards was the part complexity of task did not play in the overall calculation of scores and where the datafolio was placed on the standard setting descriptors. In hindsight - wish we had not changed the Standard Setting information on Grade 3 - ELA.
- Thanks for including me. I find this experience very helpful when working with teachers at my building with NYSAA.
- In the beginning, Grade 3, we changed the descriptors slightly. This however was not done on Grade 4. We learned that before we were very strict with the meet with distinction. However, It became easier the 2nd time around.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
English Language Arts – Grades 5 and 6**

7 Responses

(* Please note: Not all totals add up to 7 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
6	0	1	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
2	3	2	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
6	0	1

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
7	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential		
A – The Alternate Performance Level Descriptors	1	2	3	4	5
	0	0	1	4	2

	Not at all Influenced	Moderately Influenced	Very Influential		
B – The student datafolios	1	2	3	4	5
	0	2	0	4	1

	Not at all Influenced	Moderately Influenced	Very Influential		
C – Other panelists	1	2	3	4	5
	0	0	3	1	2

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	1	5	1

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
	No responses				

6. How could the standard setting process have been improved?

- Hard to work in ballrooms. Cold, big and noisy.
- The model used (student work) was not clearly outlined as to the role the work played in the process. I would suggest some initial guidance (especially for Gen. Ed. providers) on the way the work (tasks) should be used to inform thinking on rating percent
- I feel that the Standard Setting process was excellent. It provided a "true" process and open discussions.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	1	4	1

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	1	3	2

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	1	0	3	2

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	4	2

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	1	1	4	0

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	3	3

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- This process clearly brought up the issue of how well teachers are/will be trained to design an appropriate task for each child that will demonstrate their capabilities as well as the need to more clearly define and document "independence".
- Measured Progress is by far the best vendor NYS have hired in years!

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	1	4	5

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• Increased familiarity and understanding of Standard Setting tasks as day progressed.	0	0	0	1	0
• Remaining focused on what we could use to determine score.	0	0	0	0	1
• My experience – scoring and giving the Alternate Assessment.	0	0	0	0	1

6. How could the standard setting process have been improved?

- Have the student tasks weighted. It will make it easier to decide the students who are meeting the learning standards.
- AGLIs should be weighted based on level of difficulty - I think this would result in more accurate cut off points by giving us more information to consider.
- Facilitator seemed unsure of the process and at times this caused confusion with recording scores. He definitely improved by day 3.
- Instead of starting with "real" datafolios there should be a practice round to provide
- AGLIs need to be weighted or need clearer instructions as to how much we should be looking at student work in terms of deciding cut scores. There needs to be more information in the datafolios as to the type of support given. It made it very difficult to decide.
- The complexity of each AGLI should bear weight differently on the scale of scoring rubric; being clear on weight of accuracy vs. independence; level of support clearly documented.
- The first afternoon could be spent going through a mock training of grade level procedures (like the gen. ed. scoring training) so we'd be more confident of procedures when we start on the grade level datafolios.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	2	1	3	3

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	2	4	3

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	2	3	4

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	2	7

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	1	0	4	4

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	1	1	2	4

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- We also need an indication of level of support (percentage was not enough).
- I felt very comfortable with the second grade level and almost wanted to relook at first grade level. If 1st day we can run a very small mock session of the process before we actual begin looking at the ones used to set standards.
- Mr. Tracey is a wonderful facilitator.
- Richard was an excellent facilitator. Very easy going, guiding without controlling. I feel my room was well balanced and representative of the field.
- I was far more comfortable with the process during standard setting for the 2nd grad level. My confidence increased with familiarity with the process. I felt this was a very valuable process - thank you for letting me be a part of this.
- For future datafolios: Weight tasks. A check list of types of support given. First day - have a brief mock scoring session before looking at datafolios to be scored.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-11, 2008
English Language Arts – High School**

7 Responses

(* Please note: Not all totals add up to 7 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
3	4	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
1	5	1	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
5	0	2

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
5	1

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	1
			4
			2

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	2	2
			4
			2

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	1	1
			4
			1

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	1	0	1	4	1

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• The ratings/scores	0	0	0	0	1
• The guidance provided by Measured Progress staff	1	0	0	0	0

6. How could the standard setting process have been improved?

- Bigger or more diverse sample setting (datafolios) It seemed that there were way too many 4's.
- The size and make-up of the groups was excellent, the allowance for dialogue was also quite good. The process was efficient and effective. There is little I would change. The number of datafolios can skew the percentage too greatly.
- The sample of datafolios should be larger and represent a greater range of scoring.
- Datafolios should have required annotations. Datafolios should have more exemplars that are on the lower end. These datafolios appear to be "cherry picked" and don't represent "total reality".
- The number of datafolios did not appear to give a large enough variety of ranges of some continuation.
- Datafolio selection - unusually high number of datafolios that met standards.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	2	2	3	0

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	1	6	0

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	4	3

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	1	6

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	4	0	3

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	3	4

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- The process was very well constructed - efficient and thorough. However, our group questioned the datafolios representation of reality. The sample datafolios are skewed towards higher-level which is not representation of the population of kids at our schools.
- The dialogue preceding each of the steps was extremely helpful. I appreciated the opportunity to meet and discuss relevant educational topics with Assistant Commissioner Abrams. He puts a very nice, caring face on a process that is too often number driven.
- The state needs to consider datafolios for GE standards. The research and exemplars are out there.
- The cut scores reflect the nature of the assessment.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
Mathematics – Grades 3 and 4**

8 Responses

(* Please note: Not all totals add up to 8 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
5	3	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
3	4	1	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
6	0	2

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
8	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	2
			4
			1
			5

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	1
			4
			1
			6

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	6
			4
			1
			1

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	2	3	1	2

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
	No Response				

6. How could the standard setting process have been improved?

- Uniformity in scorer understanding regarding "weightedness" of tasks and task complexity issues. Isomorphic of language/reporting regarding support give.
- More standard from room to room. More Special Education panelists.
- I think we should have had lower datafolios.
- A better description of the specialized task in the invitation letter would be helpful.
- Better lighting in the room

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	0	4	4

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	0	1	7

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	1	5	2

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	1	1	1	4

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	2	5	1

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	1	6	1

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- Very interesting. Thanks for allowing me to come again. Great experience.
- There were dilemmas about looking at tasks and giving "with distinction" to students with raw scores of 48 if tasks were of lower level. Order of datafolios persuaded placing them in categories.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
Mathematics – Grades 5 and 6**

9 Responses

(* Please note: Not all totals add up to 9 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
8	1	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
0	3	6	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
6	0	3

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
9	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	1	1	2
	4	2	3

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	2
	4	0	7

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	1	1	7
	4	0	0

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	1	0	1	2	5

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• AATN and SSC experience	0	0	0	0	1

6. How could the standard setting process have been improved?

- I liked the process (stream lined) better this year than last year. Panelists I worked with were a great cross-section of expertise-better than last year.
- I feel it was very well organized, manageable and enjoyable.
- Shorter time frame. Too much down time.
- Process was well organized - no suggestions.
- Could we rate accuracy at a higher level than Independence? I didn't agree with the idea that a student could acquired zero's (receive 1pt) and score 4's on the Independence level and the final score equals 30 and that would equal a PM. Something is drastically wrong with that picture.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	3	3	3

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	2	3	4

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	0	9

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	1	1	0	7

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	0	3	6

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	2	7

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- Barbara was a great facilitator.
- Would eagerly participate in the future. I believe the process was fair - given the variables that could have influenced panelists incorrectly. Some panelists, although familiar with datafolios, did not understand the cut process - even after 4 days.
- Appreciate input from David Abrams. Input from Liz was excellent for understanding of the process. Barbara Haig was a model facilitator.
- Too bad we can't get the scores for the entire NYSAA and see where the cut lines would come out as. Barbara Haig was a supportive and helpful facilitator. Does NYS keep track of what AGLIs are used throughout the state and what level of difficulty the AGLIs come from? I would like to see that.
- This was a very interesting process. Having been an AATN and SSC this gave me a view of the next step.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
Mathematics – Grades 7 and 8**

7 Responses

(* Please note: Not all totals add up to 7 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
6	1	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
5	2	0	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
7	0	0

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
7	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	1
			4
			2

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	1
			3
			3

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	2
			2
			3

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	1	2	0	4

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• Percentages	0	0	0	0	1
• Time sequence of data entries	0	0	1	0	0

6. How could the standard setting process have been improved?

- I would have liked to receive the datafolios in random order. Feel that the fact we were told they were organized lowest to highest may have had an impact on borderline folios.
- Randomize the sample papers.
- More datafolios.
- Don't put the student datafolios in order. Allow them to be randomized. Don't skew our thinking.
- Our group would have liked the datafolio packets not be placed in any pre-order.
- Consider complexity of the task - have A,B,C,D for 4,3,2,1. Consider datafolios for other grades - like high school years.
- Set levels A,B,C re: complexity. Clearly define "prompting" for Task Administration so that mechanical prompting is clearly distinguished and eliminated from percentage independence assessment.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	1	4	2

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	1	4	2

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	1	3	3

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	1	6

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	1	5	1

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	1	4	2

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- Future Alternate Assessment formats should consider weighting scale or levels. Considering which assessment tasks are being tested. (AGLI, less complex, more complex.)
- Add level of difficulty of task chosen to the information given concerning each student's performance level.
- Level of difficulty would improve the evaluation..
- The tasks need to be grouped (leveled, tiered, etc) to aid in talking about data/info. The reason is a 4 is not a 4 is not a 4 is not a 4. When teachers share data it would be helpful to say on “A” level tasks the student is a 4 but on “B” level the student is a 3. OR in Math the student is a 3 on level “C” tasks but on ELA they are a 3 on “B” level.
- I really think the group made up of the vast experiences was wonderful. We all brought so much information to the table. I am always reminded that I (SPED) have a lot in common with Gen. Ed. I was very happy with where the percentages all fell. I think we should look at levels within our Assessment and standard rate each level of complexity 1-4 so you can show growth and weigh tasks differently.
- Content awareness: Distinguish between terms, expressions and equations and related tasks.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-11, 2008
Mathematics – High School**

7 Responses

(* Please note: Not all totals add up to 7 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
3	3	1	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
2	2	3	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
4	0	3

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
7	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	1	0	0
			4
			3
			5
			3

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	1	1
			4
			2
			5
			3

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	1	4
			4
			2
			5
			0

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	1	2	4

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
	No Responses				

6. How could the standard setting process have been improved?

- The focus was very good. I planned for 3 days. I wish that there was enough work to last for 3 days.
- More Special Education teachers on the panel would be great.
- Attention to the variation within the scores of individual datafolios and how this variation should potentially impact the overall score.
- Marking comments - rates directly on the datafolios - post-its perhaps with grids on it may reduce paper shuffling and moving datafolios from pile to pile when changing rating.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	1	0	3	2

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	1	1	0	2	3

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	3	3

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	1	1	5

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	1	1	3	2

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	2	5

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- Everything was well organized. I think that further explanation as to how cut scores are calculated could have been explained in the opening session..
- Panelists should be paid for 3 days of work regardless of when finished. For many of us we took the time off and should be compensated. The process was good but I felt there should have been more than 2 Special Education teachers with knowledge of Alternate Assessment. (on my committee)

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-11, 2008
Science – Grade 4**

8 Responses

(* Please note: Not all totals add up to 8 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
5	2	1	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
2	5	1	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
7	0	1

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
7	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	2
			2
			4

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	1
			4
			3

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	4
			2
			2

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	2	1	5

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
	No Response				

6. How could the standard setting process have been improved?

- We were scheduled for a 2.5 day time and only were here 2 days. Some people took 3 days off for this and may find it difficult to explain to the district why they were not here for the last day.
- Pay us for the days we were scheduled. Many of us took personal time off from our full time jobs and not are being shorted one day.
- Give a copy of the appropriate framework to each panelist.
- It was difficult having two groups meet in the same room. We had to be quiet when the other group was rating, even though we were in the discussion stage.
- An overview of the Alternate Assessment process should be provided for general educators.
- I don't think anything can be improved as I have learned a lot about others, met great peers and co-workers from across the state. I received some great input and grew as a teacher from this. I am thankful for this opportunity and enjoyed meeting staff from the NYS Education Department.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	1	4	3

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	0	3	5

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	5	3

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	2	6

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	2	2	4

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	1	5	2

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- In the Performance Level Descriptors it states that if a student is partially meeting learning standards, they need maximum support. This should only moderate or medium support as if the student can do some of the task requested - they will not need the full maximum support in order to finish.
- This is the second year in a row that we were sent home early.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
Science – Grades 8 and High School**

7 Responses

(* Please note: Not all totals add up to 7 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
5	2	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
0	5	2	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
7	0	0

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
7	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	1
			4
			2
			4

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	1
			3
			3

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	2
			4
			1

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	1	1	5

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• Data results	0	0	0	1	0
• Discussion	0	0	0	1	0

6. How could the standard setting process have been improved?

- While I enjoyed my group, our experience was very similar (Special Education students with severe disabilities, mostly rural areas). I think feed-back from a General Education teacher would have "challenged" our decisions more.
- What a marvelous job! Our group was great however we could have used more diversity.
- Our group needed to be more diverse. We had no General Education teachers. Our group worked well together and was able to "argue" efficiently and effectively, but was under-represented geographically and background-wise.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	2	0	3	2

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	1	3	2

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	1	3	3

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	0	0	7

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	0	1	6

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	4	3

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- This was a super experience. I will go back to my agency with greater insight about this process. I will jump at the chance to do this again. Thank you!
- Wonderful facilitator!
- I think some more information about the Psychometricians and this process - perhaps a bit every morning - to broaden our general understanding of the process we are a part of. Luz explained everything very well and piqued our interest in the process and how the data is used.
- David Abrams is a very good speaker. Very interesting.
- I was very excited to see the impact data. I understand the process much better now.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-11, 2008
Social Studies – Grade 5**

8 Responses

(* Please note: Not all totals add up to 8 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
6	0	2	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
4	3	1	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
7	0	1

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
8	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	1
			4
			3
			4

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	0
			4
			3
			5

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	6
			4
			1
			1

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	0	0	0	6	2

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
• Clarifications/input from facilitator	0	0	0	1	0
• Level of work, AGLI, tasks	0	0	0	1	0

6. How could the standard setting process have been improved?

- I thought it was extremely well organized and efficient. Sharon was a superior facilitator.
- Better training for teachers giving the NYSAA (datafolio)
- Testing needs to be developed several per topic instead of teacher made test.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	2	0	2	0	4

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	0	0	3	3	2

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	1	3	4

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	0	1	4	3

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	2	4	2

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	4	4

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- The Standard Setting Quick Reference for NYSAA is a great document. Is there any way to share this with new AATNs and SSCs? Thank you for a great experience. I learned so much.
- I feel the complexity score should be added to datafolios. This would assist in differentiating between the meeting and meeting with distinction students.
- Sharon was great! Thanks for this experience.

**NYSAA Standard Setting
Evaluation Response Summary
Albany, June 9-13, 2008
Social Studies – Grades 8 and High School**

8 Responses

(* Please note: Not all totals add up to 8 as not everyone responded to all questions.)

1. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

Very Good	Good	Neutral	Poor	Very Poor
4	4	0	0	0

2. How clear were the Alternate Performance Level Descriptors?

Very Clear	Clear	Somewhat Clear	Not Clear
2	5	1	0

3. How would you judge the length of time of this meeting for setting alternate performance standards?

About Right	Too Little Time	Too Much Time
8	0	0

4. Do you believe the cut scores resulting from the panel's Round 3 ratings are correctly placed?

Yes	No
8	0

5. Which is your overall impression of the process used to set performance standards for the New York State Alternate Assessment?

	Not at all Influenced	Moderately Influenced	Very Influential
A – The Alternate Performance Level Descriptors	1	2	3
	0	0	1
			4
			5
			2

	Not at all Influenced	Moderately Influenced	Very Influential
B – The student datafolios	1	2	3
	0	0	0
			4
			4
			4

	Not at all Influenced	Moderately Influenced	Very Influential
C – Other panelists	1	2	3
	0	0	1
			4
			2
			5

	Not at all Influenced	Moderately Influenced	Very Influential		
D – My experience in the field	1	2	3	4	5
	1	0	1	3	3

	Not at all Influenced	Moderately Influenced	Very Influential		
E – Other (please specify)	1	2	3	4	5
	No Response				

6. How could the standard setting process have been improved?

- More clarification on the Alternate Performance Level Descriptors (i.e.: rarely; often - what do they mean?)
- Could use a snack at 2:30 to re-energize. Group really started to fade in late afternoon.
- Clarify "often/rarely" on descriptors.
- Keep groups diversified. Have facilitator who is knowledgeable about the NYSAA in all of its complexity.
- Make sure questions/prompts are clear.
- Keep groups diversified. Have facilitator who is knowledgeable about the NYSAA in all of its complexity.
- Make sure questions/prompts are clear.
- More panelists from different racial/ethnic backgrounds to offer a different point of view/experience.
- One way to improve the process would be to weight the complexity of the tasks/AGLIs. Since they are already broken up into degrees of complexity the value they hold must have some influence in scoring. This is the only way to determine if students with disabilities can meet higher standards. Teachers then would have to push their students that were capable to reach for higher standards without being penalized if they received a lower score.

For each statement below, please circle the rating that best represents your judgment.

	Not at all Useful		Very Useful		
7. The opening session was:	1	2	3	4	5
	0	0	1	1	6

	Not at all Useful		Very Useful		
8. Providing additional details to the Alternate Performance Level Descriptors was:	1	2	3	4	5
	1	0	1	1	5

	Not at all Useful		Very Useful		
9. The datafolio rating task was:	1	2	3	4	5
	0	0	0	4	4

	Not at all Useful		Very Useful		
10. The discussion with other panelists was:	1	2	3	4	5
	0	1	0	1	6

	Not at all Useful		Very Useful		
11. The impact data provided was:	1	2	3	4	5
	0	0	2	2	4

	Not at all Useful		Very Useful		
12. I was confident in classifying the student datafolios:	1	2	3	4	5
	0	0	0	3	5

Additional Comments

13. Please provide any additional comments or suggestions about the standard setting process.

- Good group dynamic. Moderator kept us on task. Very informative. Great job.
- Five days was not really needed. This entire assessment could have been done in no more than four days. A "hot" breakfast should have been provided. The lunch was very good.
- I felt the breakfast was not very good. Since people will be working all day a better breakfast would provide participants with additional fuel for this task. Lunch could have had more choices.
- Should prompts be allowed or not. Need specific clarification. Should AGLIs be weighted? Should students who are medically fragile be penalized because they are doing less complex tasks and scoring 4's in accuracy and independence? Very difficult to remove administration and scoring caps and experience. Different groups told different things about complexity - same message to all groups.
- This was a great experience working with the standard setting panel. It was a pleasure in understanding the difference of districts that sat on the panel.
- If the data collection sheets are going to be continued to be used, then they should be scored in a different manner than datafolios that contain actual student work. Maybe there should be levels of distinction, similar to levels of distinction in General Education would. (i.e.: honors and honors-plus. For the most significantly impaired students, there should be a separate assessment tool or instruments. Is this an attempt to standardize a non-standard test?
- There needs to be clarification in regard to what constitutes meeting with distinction, especially when there is no evidence present and the teacher's description of the tasks is unclear.

APPENDIX I—POLICY FORUM AGENDA

Agenda and Process Overview
Policy Forum Discussion
Standard Setting
Friday, June 13, 2008
9am to 11am
Crowne Plaza
Albany, NY

8:30 – 9:00	Breakfast in Pearl Street Room	
9:00 – 9:10	Welcome and Introductions	NYSED
9:10 – 9:20	Purpose, Process and Outcomes	MP
9:20 - 9:35	Review Materials All grades/all content areas	Luz Bay
9:35 – 10:50	Group Discussions and Recommendations English Language Arts Mathematics Science Social Studies Evaluations	Luz Bay
10:50 – 11:00	Wrap up	NYSED/MP

Purpose: To elicit recommendations on final performance level cuts from expert stakeholder group.

Process: Following breakfast, the whole group will meet. The Department will give a welcome, thank you & introductions. Measured Progress will provide an overview of the morning’s activities:

1. Review standard setting purpose & how this activity supports standard setting
2. Set the stage for the group discussion
3. Hand out pertinent materials
 - a. APLDs
 - b. Scoring Rubric
 - c. Results from 2006-2007 standard setting
 - d. Raw score distribution from 2007-2008
 - e. Round 3 results from 2007-2008 standard setting
4. Present Round 3 standard setting results
5. For each subject, panelists will be asked if the standard setting results are deemed reasonable. They will engage in discussion regarding the reasonableness of the results and arrive at a recommendation whether to modify the cutpoints.
6. If the panel recommends modification, a discussion to arrive at new recommended cutpoints will ensue. The effect of the changes will be shown to the panel in real time.
7. Participants complete an evaluation form.
8. The Department will close the meeting.

Outcome: Recommendations for final performance level cut points.

***APPENDIX J—POLICY FORUM POWERPOINT
PRESENTATION***

Slide 1

**New York Student
Alternate Assessment**

Standard Setting
Policy Forum

 University of the
State of New York
State Education
Department

 Measured
Progress

Slide 2

Purpose of Standard Setting

To establish cut scores that distinguish between New York's four alternate performance levels

- *Meets the Learning Standards with Distinction*
- *Meets the Learning Standards*
- *Partially Meets the Learning Standards*
- *Does Not Meet the Learning Standards*

2

Slide 3

Purpose of Policy Forum

To elicit recommendations on the final performance level cut scores from expert stakeholder group.

3

Slide 4

Charge to the Committee

- Examine the cut scores and impact data
- Determine if the cut scores are reasonable
- If necessary, modify one or more of the cut scores to make them reasonable

4

Slide 5

Things to Consider

- Alternate Performance Level Descriptors (APLDs)
- Distribution of students by grade, content area, and alternate performance level
- Raw scores

5

Slide 6

Chief Ground Rule

Any modification made to any cut score would have to be grounded in the content of the respective APLDs. Thus, it will not be permissible to raise or lower a cut score simply to smooth a line. The new cut score would have to be justified on the basis of the content area assessment and the APLDs. (Cizek & Bunch, 2007, p. 267)

6

Slide 7

Materials

- Agenda
- Alternate Performance Levels Descriptors (APLDs)
- Scoring Rubric
- [Round 3 Results from 2007-2008 Standard Setting](#)
- [Raw Score Distributions](#)
- [Results from 2006-2007 Standard Setting](#)
- Evaluation Form

7

Slide 8

Round 3 Results

- [English Language Arts](#)
- [Mathematics](#)
- [Science](#)
- [Social Studies](#)

8

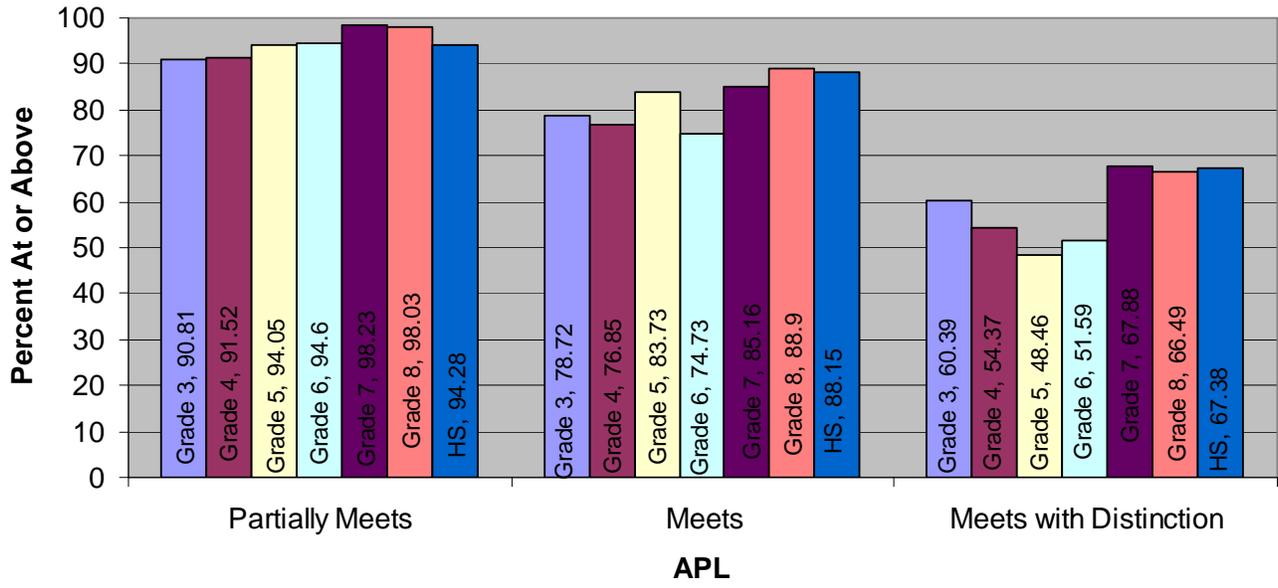
Slide 9

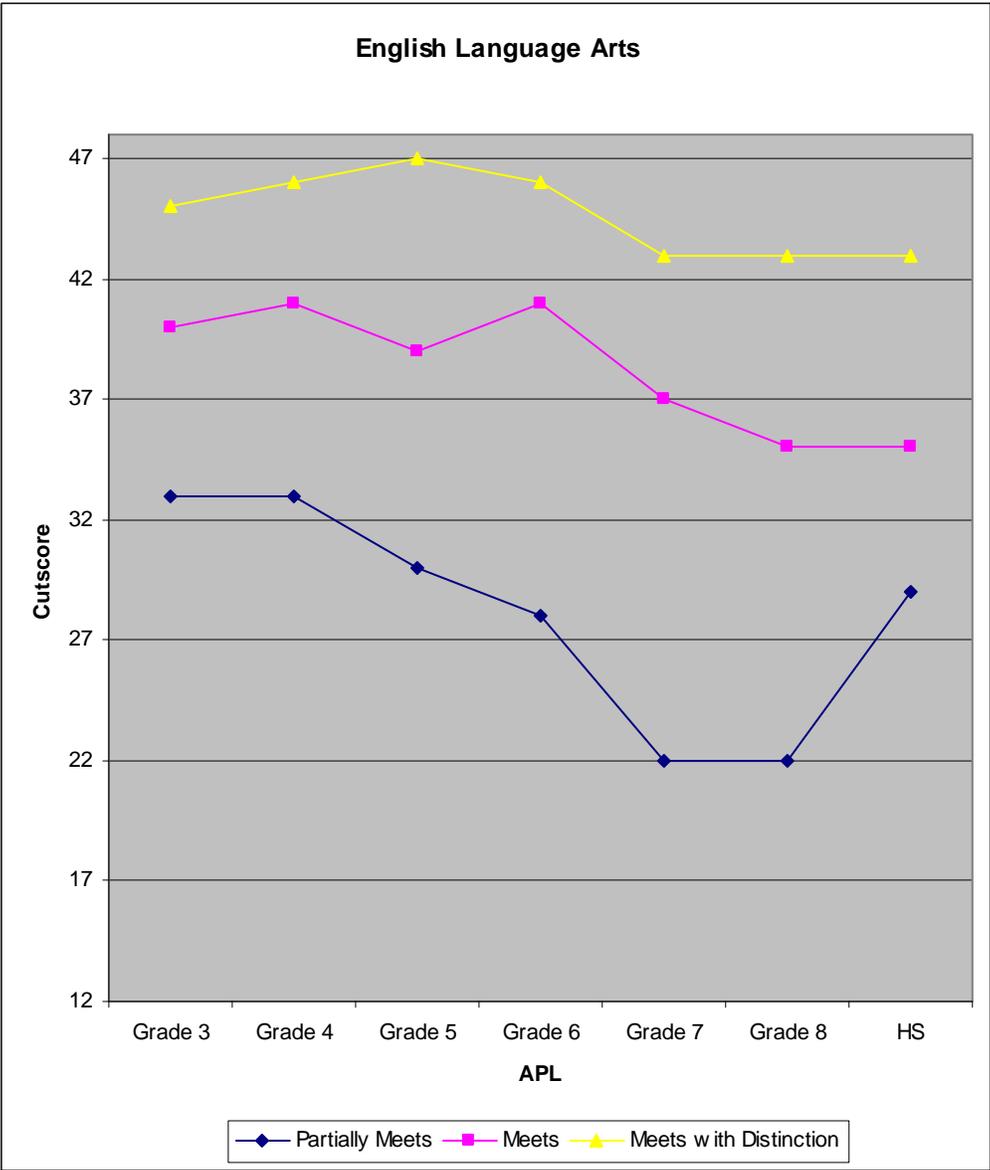
Thank you for your participation!

APPENDIX K—POLICY FORUM TABLES AND GRAPHS

English Language Arts									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3	33	12.09	90.81	40	18.33	78.72	45	60.39	60.39
Grade 4	33	14.67	91.52	41	22.48	76.85	46	54.37	54.37
Grade 5	30	10.32	94.05	39	35.27	83.73	47	48.46	48.46
Grade 6	28	19.87	94.6	41	23.14	74.73	46	51.59	51.59
Grade 7	22	13.07	98.23	37	17.28	85.16	43	67.88	67.88
Grade 8	22	9.13	98.03	35	22.41	88.9	43	66.49	66.49
High School	29	6.13	94.28	35	20.77	88.15	43	67.38	67.38

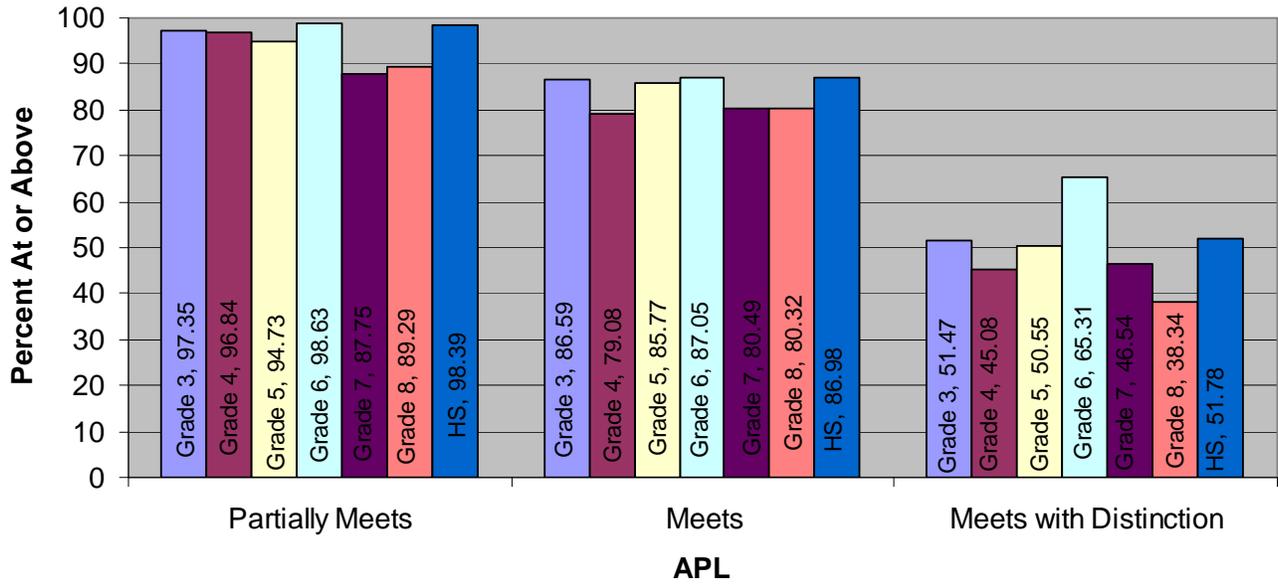
English Language Arts

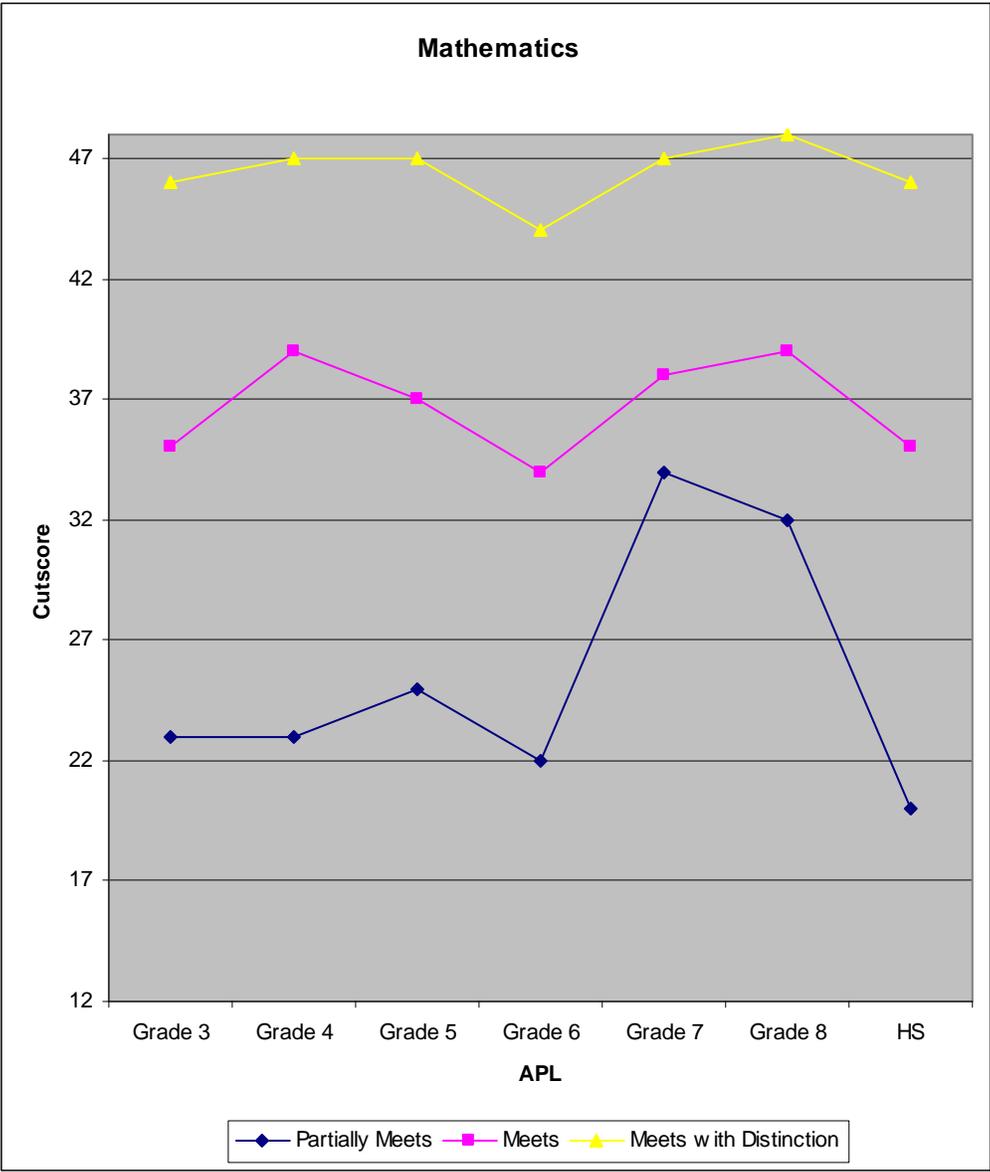




Mathematics									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3	23	10.76	97.35	35	35.12	86.59	46	51.47	51.47
Grade 4	23	17.76	96.84	39	34	79.08	47	45.08	45.08
Grade 5	25	8.96	94.73	37	35.22	85.77	47	50.55	50.55
Grade 6	22	11.58	98.63	34	21.74	87.05	44	65.31	65.31
Grade 7	34	7.26	87.75	38	33.95	80.49	47	46.54	46.54
Grade 8	32	8.97	89.29	39	41.98	80.32	48	38.34	38.34
High School	20	11.41	98.39	35	35.2	86.98	46	51.78	51.78

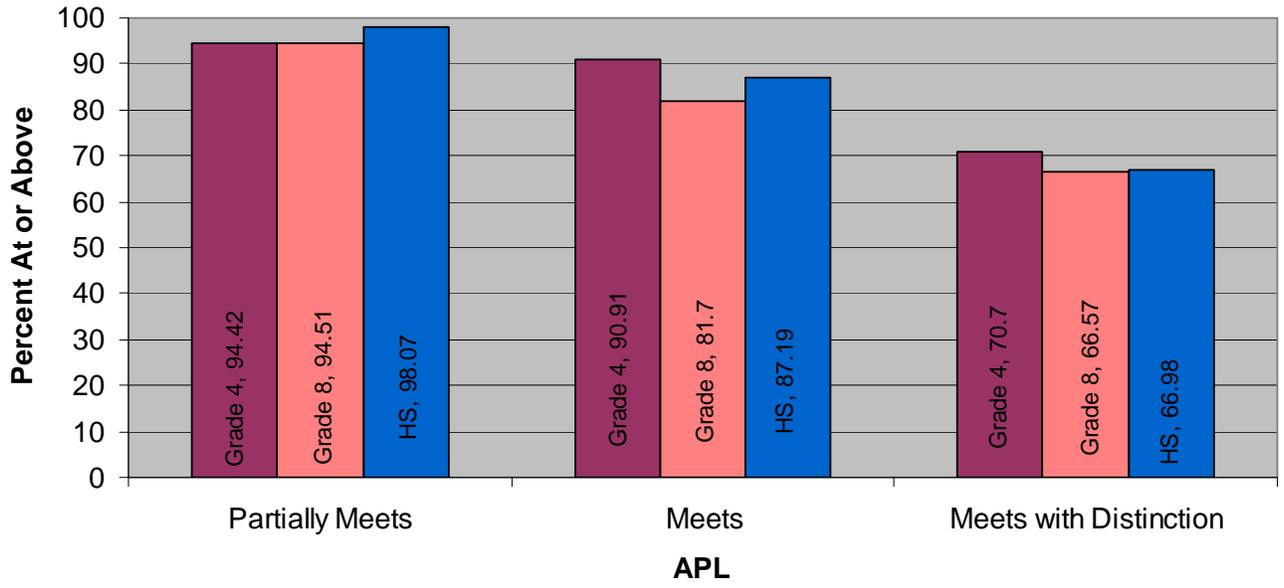
Mathematics

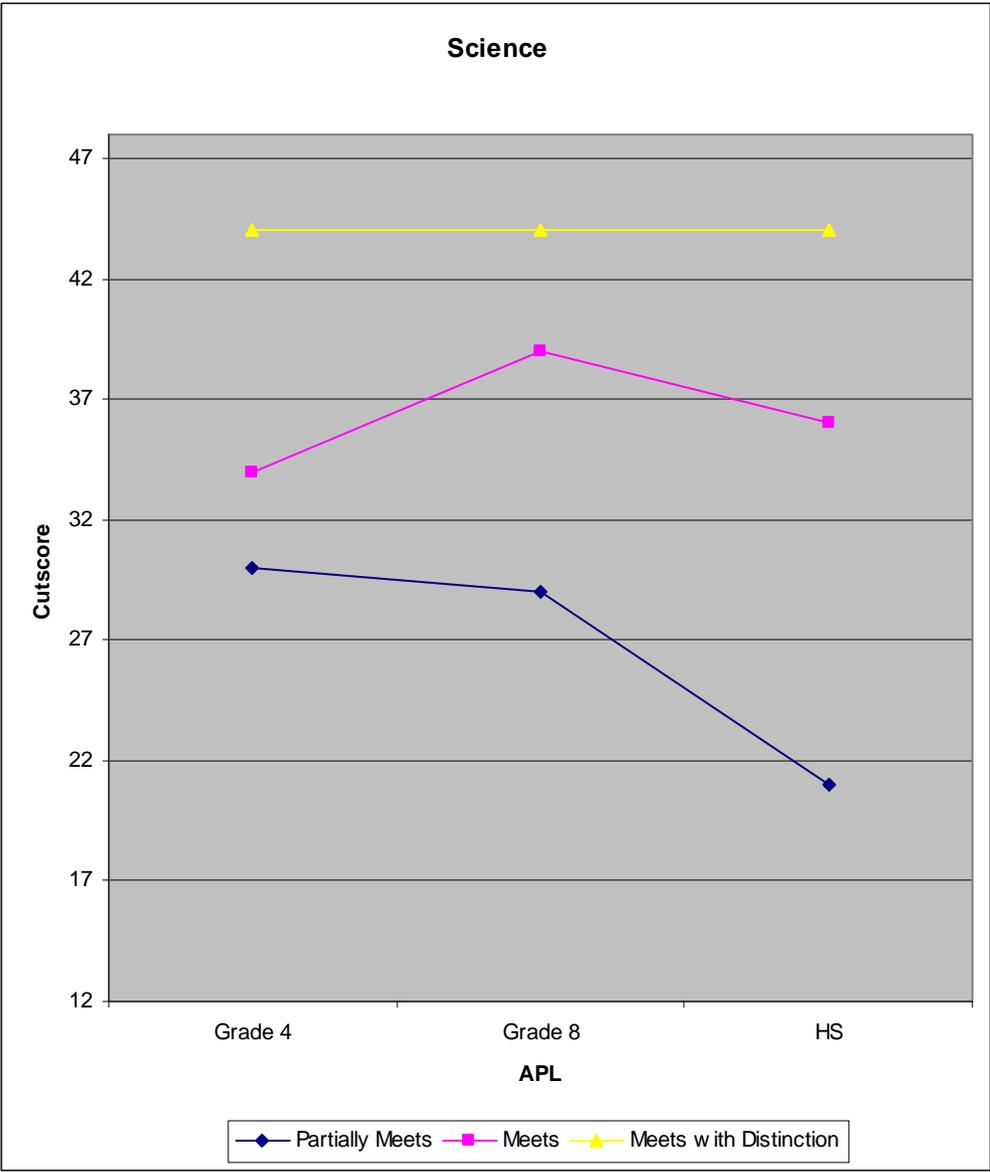




Science									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3									
Grade 4	30	3.51	94.42	34	20.21	90.91	44	70.7	70.7
Grade 5									
Grade 6									
Grade 7									
Grade 8	29	12.81	94.51	39	15.13	81.7	44	66.57	66.57
High School	21	10.88	98.07	36	20.21	87.19	44	66.98	66.98

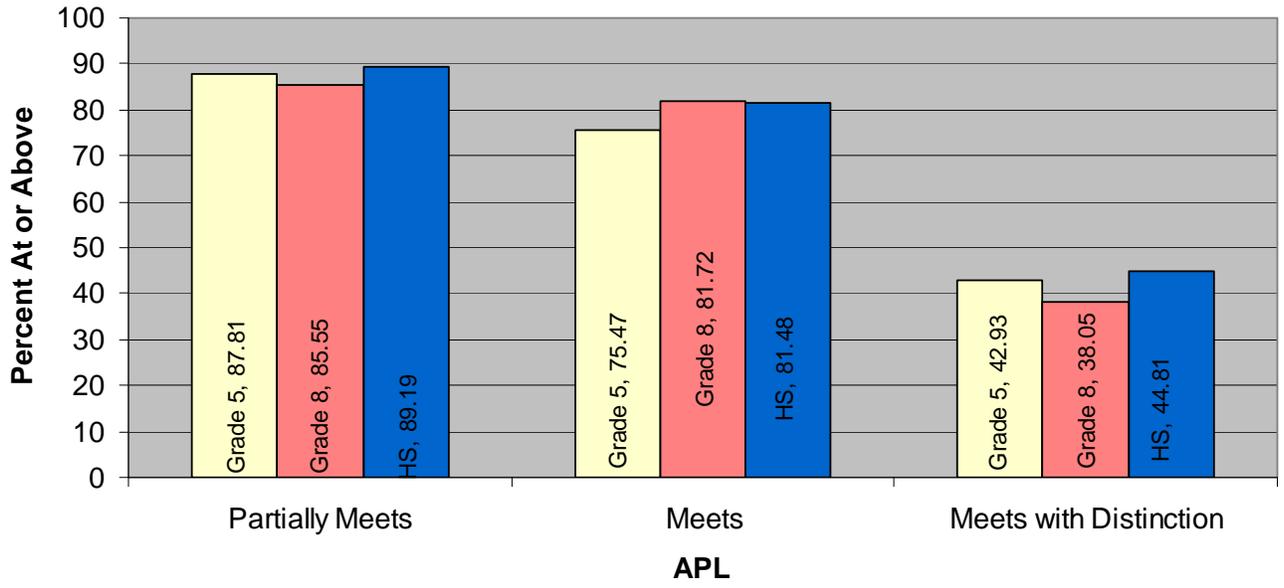
Science

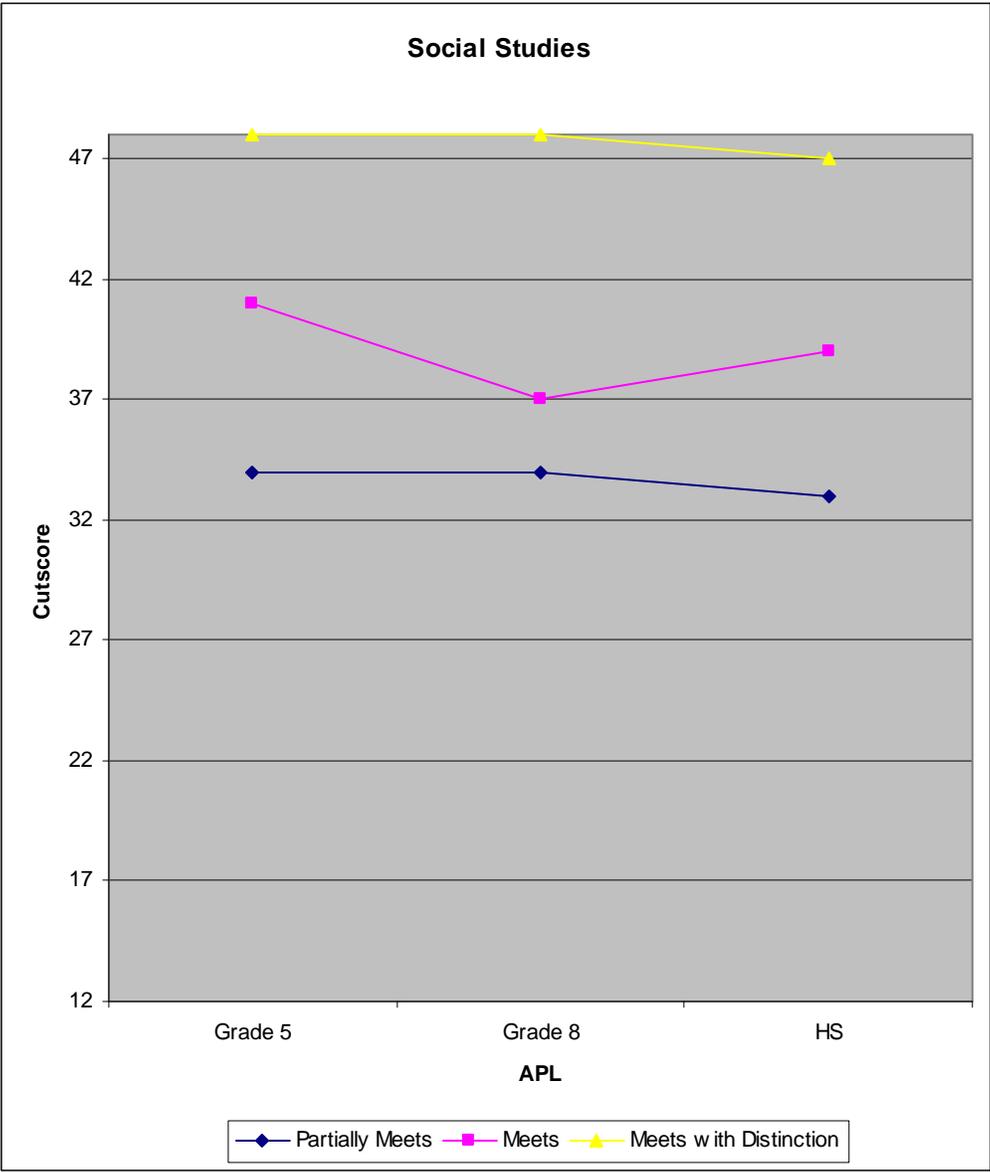




Social Studies									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3									
Grade 4									
Grade 5	34	12.34	87.81	41	32.54	75.47	48	42.93	42.93
Grade 6									
Grade 7									
Grade 8	34	3.83	85.55	37	43.67	81.72	48	38.05	38.05
High School	33	7.71	89.19	39	36.67	81.48	47	44.81	44.81

Social Studies



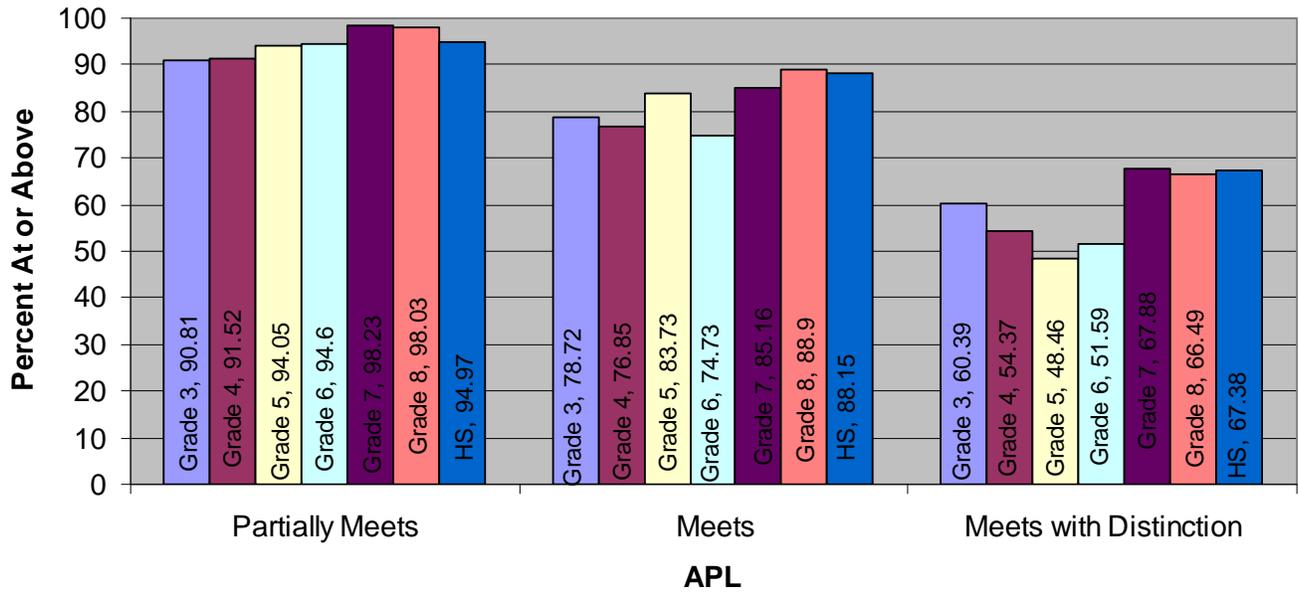


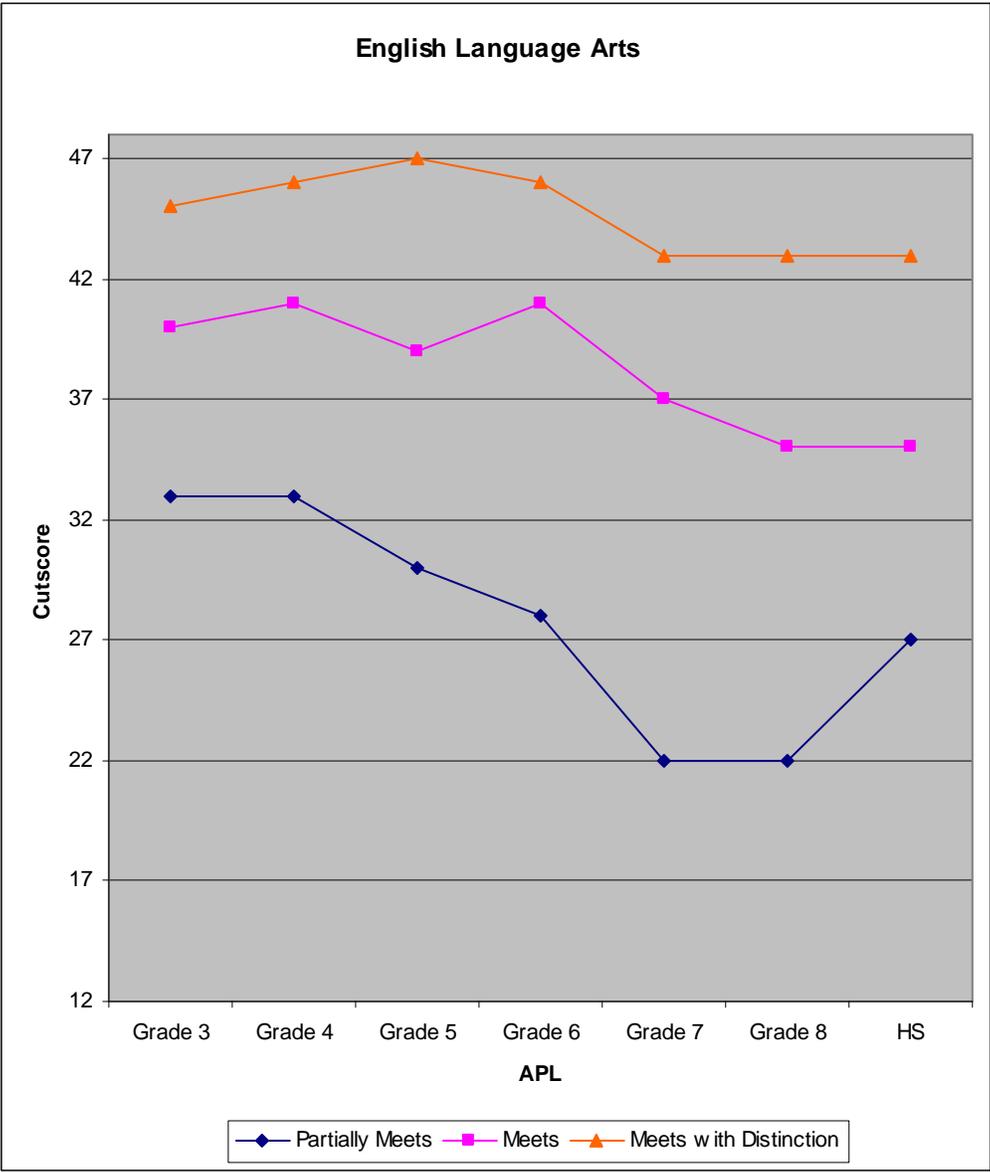
APPENDIX L—POLICY FORUM TABLES AND GRAPHS OF RECOMMENDATIONS

English Language Arts

Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3	33	12.09	90.81	40	18.33	78.72	45	60.39	60.39
Grade 4	33	14.67	91.52	41	22.48	76.85	46	54.37	54.37
Grade 5	30	10.32	94.05	39	35.27	83.73	47	48.46	48.46
Grade 6	28	19.87	94.6	41	23.14	74.73	46	51.59	51.59
Grade 7	22	13.07	98.23	37	17.28	85.16	43	67.88	67.88
Grade 8	22	9.13	98.03	35	22.41	88.9	43	66.49	66.49
High School	27	6.82	94.97	35	20.77	88.15	43	67.38	67.38

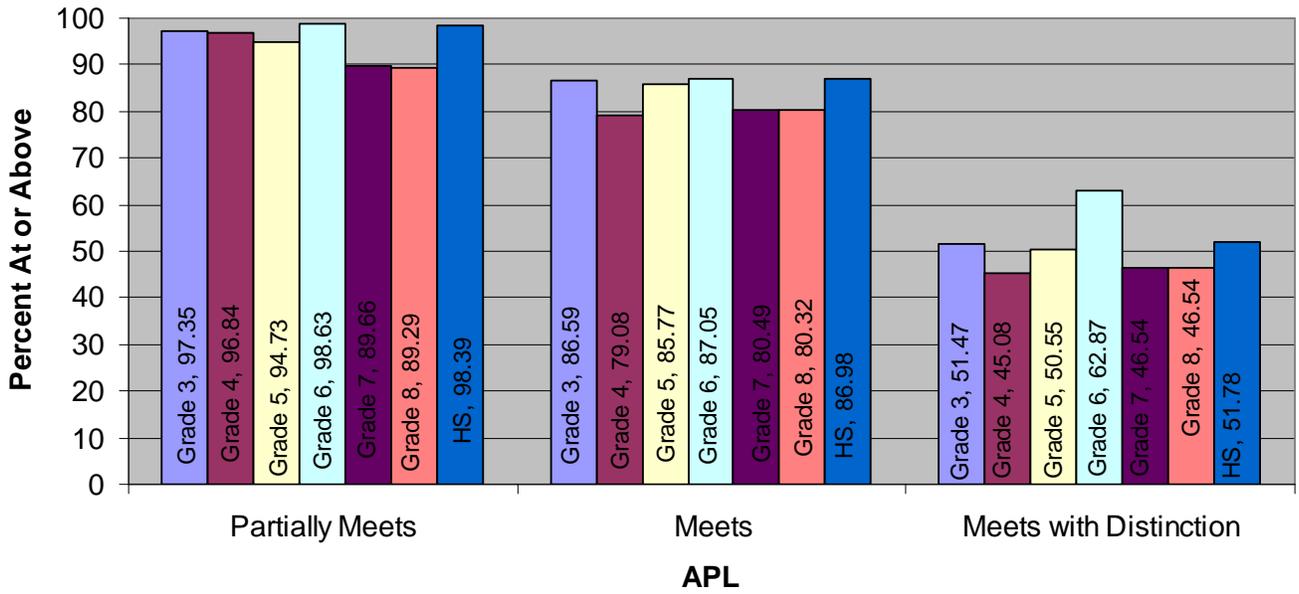
English Language Arts

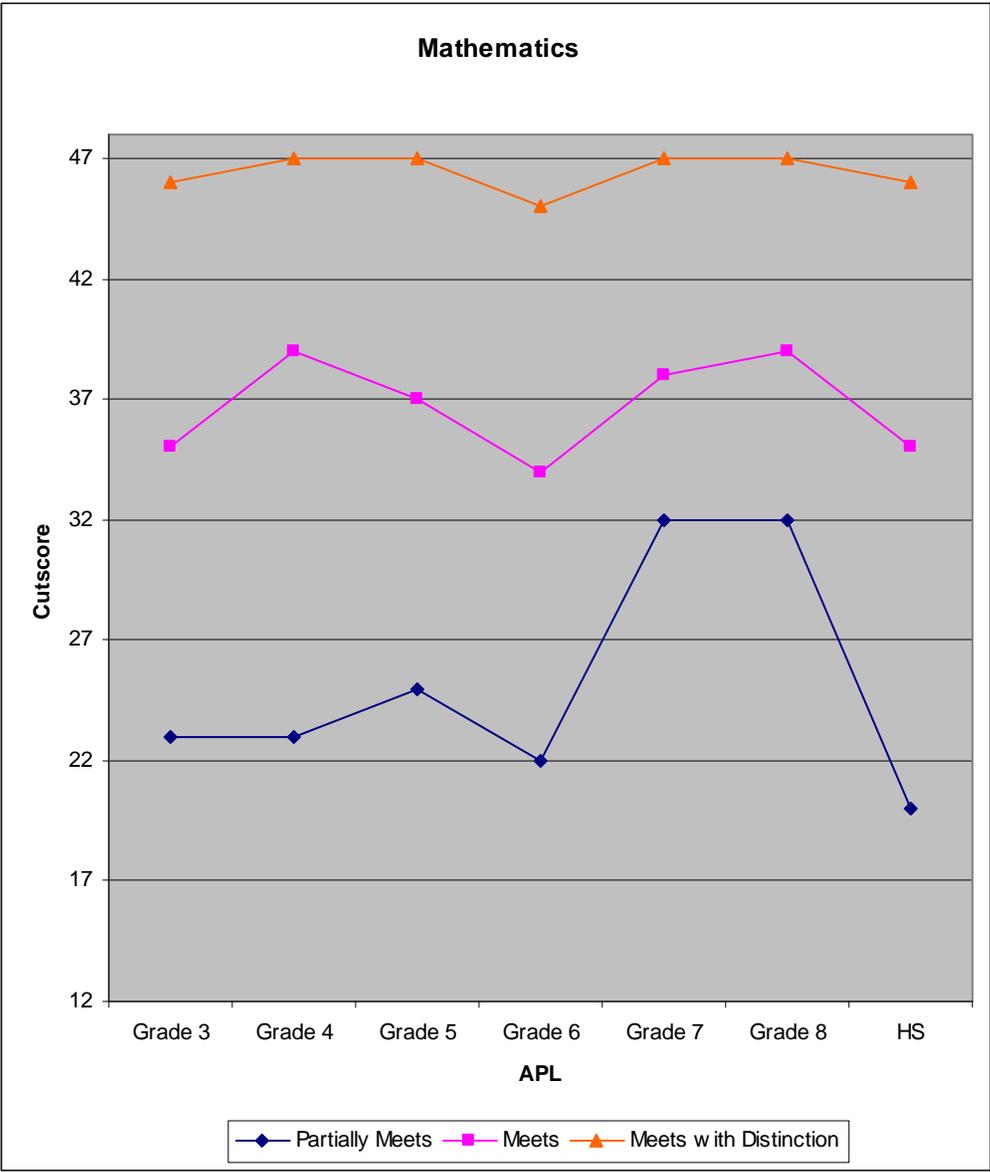




Mathematics									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3	23	10.76	97.35	35	35.12	86.59	46	51.47	51.47
Grade 4	23	17.76	96.84	39	34	79.08	47	45.08	45.08
Grade 5	25	8.96	94.73	37	35.22	85.77	47	50.55	50.55
Grade 6	22	11.58	98.63	34	24.18	87.05	45	62.87	62.87
Grade 7	32	9.17	89.66	38	33.95	80.49	47	46.54	46.54
Grade 8	32	8.97	89.29	39	33.78	80.32	47	46.54	46.54
High School	20	11.41	98.39	35	35.2	86.98	46	51.78	51.78

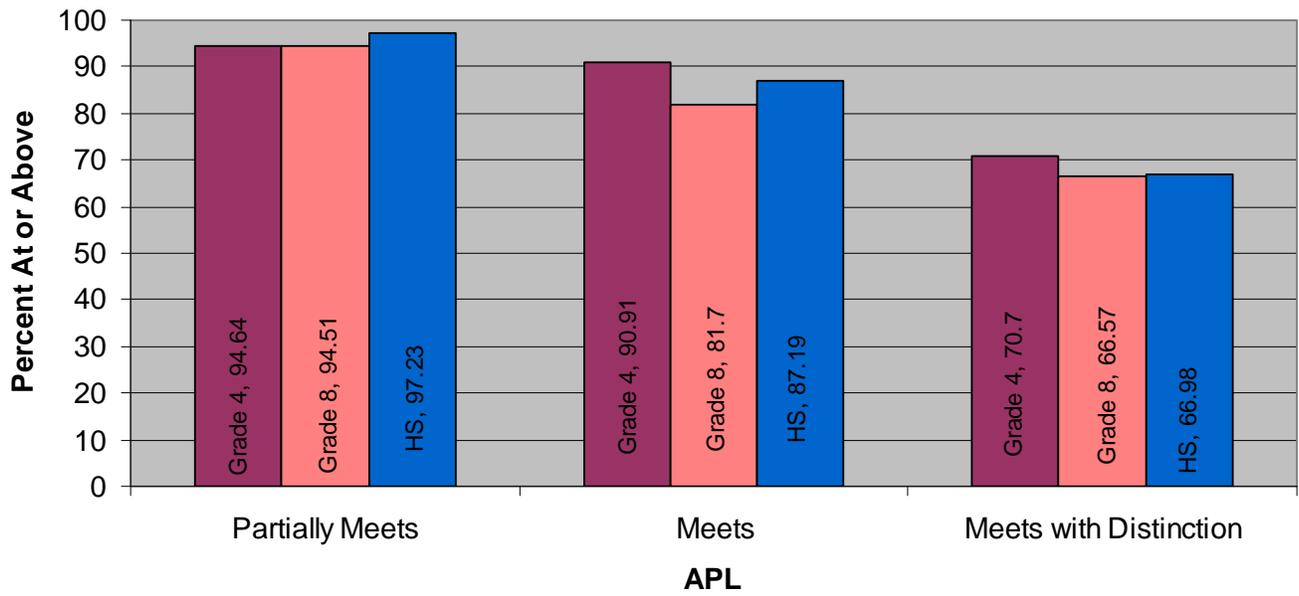
Mathematics

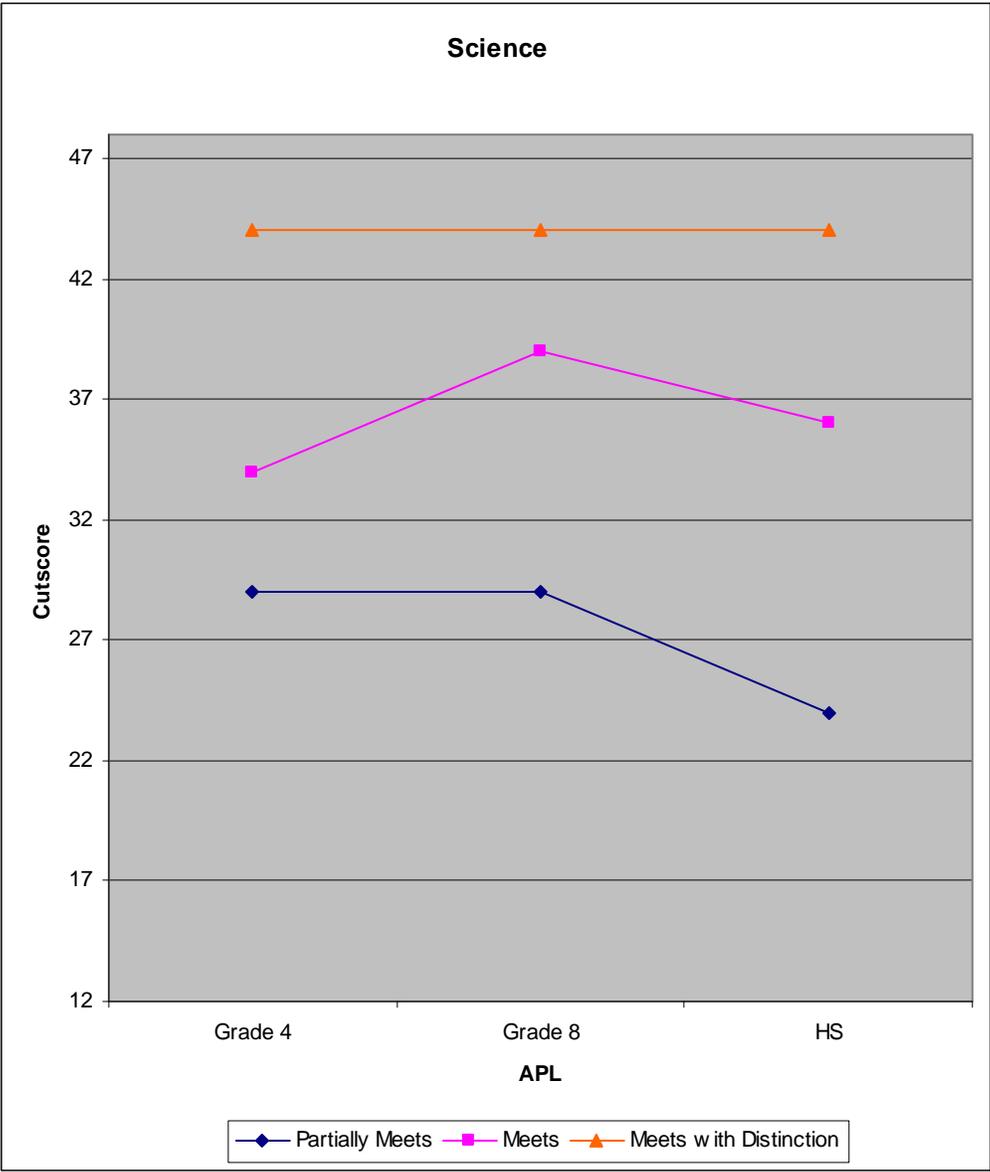




Science									
Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3									
Grade 4	29	3.73	94.64	34	20.21	90.91	44	70.7	70.7
Grade 5									
Grade 6									
Grade 7									
Grade 8	29	12.81	94.51	39	15.13	81.7	44	66.57	66.57
High School	24	10.04	97.23	36	20.21	87.19	44	66.98	66.98

Science

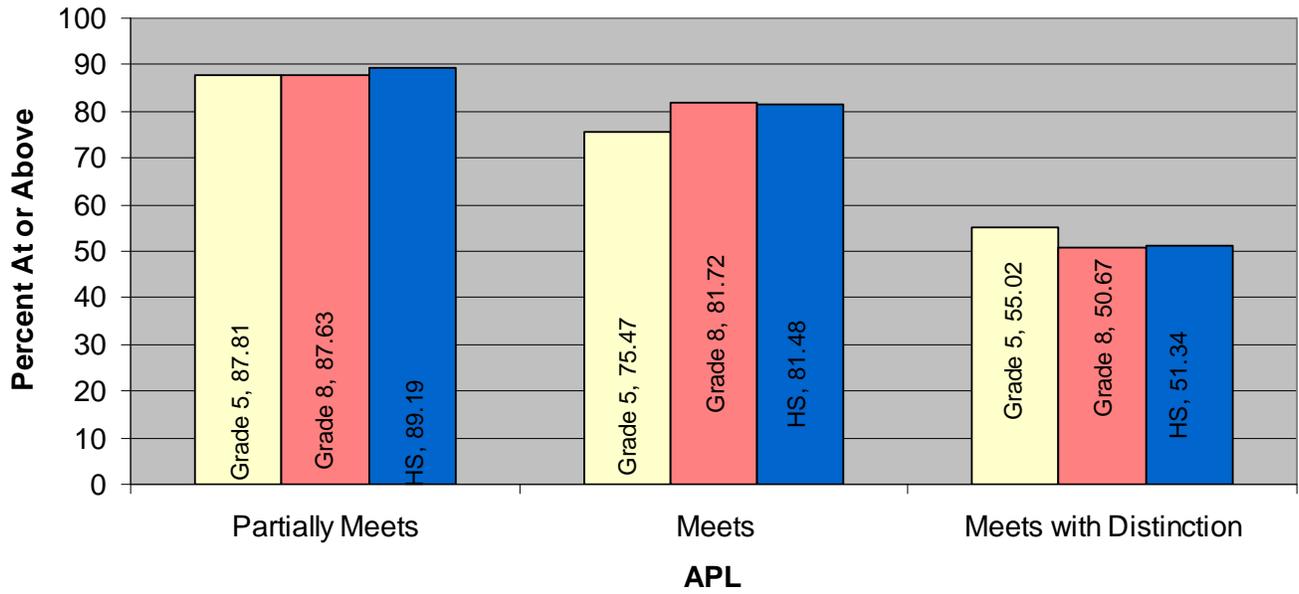




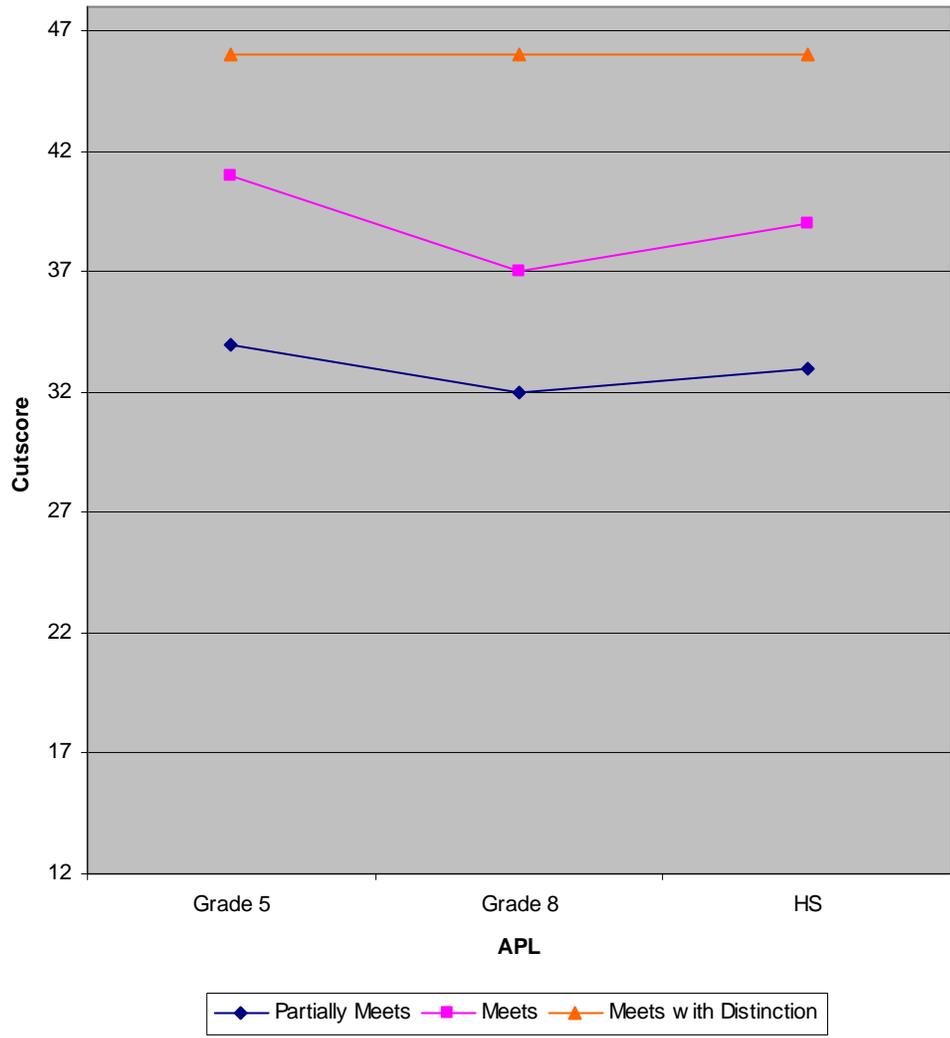
Social Studies

Grade	Partially Meets			Meets			Meets with Distinction		
	Cut 1	%	%=>	Cut 2	%	%=>	Cut 3	%	%=>
Grade 3									
Grade 4									
Grade 5	34	12.34	87.81	41	20.45	75.47	46	55.02	55.02
Grade 6									
Grade 7									
Grade 8	32	5.91	87.63	37	31.05	81.72	46	50.67	50.67
High School	33	7.71	89.19	39	30.14	81.48	46	51.34	51.34

Social Studies



Social Studies



***APPENDIX M—SAMPLE EVALUATION FORM AND RESULTS
OF POLICY FORUM EVALUATION BY PANELISTS***

Policy Forum Evaluation Form

Name: _____

ID: _____

Item	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
1	The purpose of the policy forum was clearly explained.				
2	The tasks assigned to the panel were explained sufficiently.				
3	The APLDs were clear and useful.				
4	There was adequate time provided for discussions.				
5	There was an equal opportunity for everyone in the group to contribute his/her ideas or opinions.				
6	I am confident about the defensibility and appropriateness of the final recommended cut scores.				
7	The recommended cut scores represent a realistic correspondence between student performance and the APLDs.				

Comments:

Standard Setting Policy Forum
June 13, 2008 - Albany Crowne Plaza
13 responses

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
The purpose of the policy forum was clearly explained.	11	2	0	0
The tasks assigned to the panel were explained sufficiently.	12	1	0	0
The APLDs were clear and useful.	6	7	0	0
There was adequate time provided for discussions.	12	1	0	0
There was an equal opportunity for everyone in the group to contribute his/her ideas or opinions.	12	1	0	0
I am confident about the defensibility and appropriateness of the final recommended cut scores.	11	2	0	0
The recommendation of cut scores represent a realist correspondence between student performance and the APLDs.	12	1	0	0

Comments

I see a great deal of usefulness for this process/information to be shared with AATNs and LSETs. Thanks you for allowing me to be included with this.
I am curious about the raw scores at 18, 24, 30, 36, 42 and possibly 48. If investigated, could possible reasons be shared?
The two scores (accuracy and independence) imply equivalence, but the descriptors only appear to address it is on of the bullets. This may have created confusion on questions that were compounded by the issue of complexity.
No weight given to AGLIs but for purposes of talking with teachers, level the tasks. "A" level with an "A-4", etc...
Please consider establishing level for the tasks for each AGLI. It will make standard setting easier and more importantly it will provide a better picture of the ability levels of the students.
This too longer than I had anticipated though I fell that our results were useful, defensible, and valid. At times I did feel that some panelists lost sight of the nature and severity of the disabilities of some of the student who are administered NYSAA.
The process was fascinating. I have learned a lot and will use this knowledge in training my teachers on the assessments. I would be very interested in participating again. The scheduling for this panel was not realistic - morning over 1:45pm.
This was a great experience that I will apply when administering the NYSAA. Thanks for having me.
The APLDs need refinement.