New York State
Regents Examination in Geometry

Standard Setting Technical Report

Prepared for the New York State Education Department
by Pearson
September 18, 2009

Table of Contents

List of Tables

List of Tables (Continued)

List of Figures

Executive Summary

The standard setting process for the Regents Examination in Geometry consisted of three activities: the Pre-Policy Measurement Review Panel meeting, the Item Mapping Standard Setting meeting, and the Post-Policy Measurement Review Panel meeting. This document provides a detailed description of each of these activities. The main purpose of these standard setting activities was to obtain cut score recommendations for the New York State Regents Examination in Geometry. Students could be classified into the following three achievement levels on the assessment: the lowest level, 0–64 (Level 1); 65–84 (Level 2); and the highest level, 85–100 (Level 3).

On Thursday, April 9, 2009, a Pre-Policy Measurement Review meeting was conducted in Albany, New York. This meeting was convened to provide recommendations for the acceptable percentage of New York State students who should be classified in each achievement level on the New York State Regents Examination in Geometry.

On Monday, June 22, and Tuesday, June 23, an item mapping standard setting meeting was conducted using two committees. During the afternoon of Tuesday, June 23, selected members of the two committees also formed a synthesis group to reconcile the recommendations from the two independent committees. The purpose of this meeting was to recommend cut scores based on the content standards and achievement level descriptors for the same assessment.

Finally, during the evening of Tuesday, June 23, a Post-Policy Measurement Review Panel meeting was conducted. This meeting, which included panelists from the Pre-Policy Measurement Review Panel, integrated results from the Pre-Policy Measurement Review Panel meeting and the item mapping standard setting meeting.

In this technical report, panelists, materials, methodologies, and results are presented for each of the three stages for the standard setting activity for the New York State Regents Examination in Geometry. A separate executive summary was provided to the state the day after the standard setting activity outlining the methodologies and major findings. More details are provided in the current technical report.

Pre-Policy Measurement Review Panel

On Thursday, April 9, 2009, the Pre-Policy Measurement Review Panel meeting was held in Albany. This meeting was convened to provide recommendations for the percentage of New York State students who should be classified into each achievement level on the Regents Examination in Geometry. During this meeting, panelists participated in two rounds of discussions in which they were asked to make individual "high" and "low" recommendations as to the percentage of students who should be classified into each achievement level. For example, a committee member could recommend that 25–30% of students be classified as Level 3.

Panelists

A total of 30 panelists attended the meeting. These panelists were policyholders and administrators who were geographically representative of New York State. The panelists represented various stakeholder groups such as School Administrators Association of New York State (SAANYS), New York State United Teachers (NYSUT), New York State Council of School Superintendents (NYSCOSS), New York State School Board Association (NYSSBA), Big Five Cities, special education directors, district superintendents, assistant superintendents, superintendents of schools, etc. All panelists provided voluntary demographic information. Out of the 30 panelists, 18 were female and 12 were male. In terms of demographics, 3 panelists were African American, 1 was Hispanic, 23 were white, 1 was of another ethnicity, and 2 did not complete this response. In terms of geographic locations, 4 panelists were from north New York State, 10 were from central New York State, 4 were from east New York State, 5 were from west New York State, 6 were from south New York State, and 1 panelist did not complete this response.

Method and Procedure

The Pre-Policy Measurement Review Panel meeting was held on Thursday, April 9. The agenda for the meeting is shown in Appendix A. The meeting began with a review of the agenda and introductions of the facilitators and New York State Education Department (NYSED) staff. A description of standard setting was then presented, and panelists were introduced to the purpose of the meeting and the role they would play in the process. Next, Pearson facilitators described the procedure that would be used for the meeting. Panelists were told that the outcome of this meeting would be their recommendations for an acceptable range of impact data for the Regents Examination in Geometry.

After the midmorning break, panelists were presented with a set of five questions. Panelists were instructed that these questions should guide their

consideration of the test data they would subsequently see. The five questions were:

1. What type of differences in impact data do the participants expect across performance levels?
    - Equal across performance levels?
    - Increasing across performance levels?
    - Decreasing across performance levels?

2. What percentage of students in each performance level would the panel find acceptable on the new examination?
    - What would be the ideal percentage of students in each performance level?
    - What variations from these ideal values are acceptable?

3. What, if any, consistency is expected between the data from the current and new testing programs?
    - Should the percentage of students in each performance level be similar, even if the standards have changed?
    - What differences in impact data between the current and new testing programs are acceptable?

4. What type of consistency in impact data does the panel expect among the New York State Grade 8 Mathematics Test, the New York State Regents Examination in Integrated Algebra, and the New York State Regents Examination in Geometry?
    - What are the differences in impact data among the three testing programs?
    - Should the percentage of students in each performance level be similar, even though the tests measure different knowledge and skills?
    - What differences among the three programs are acceptable?

5. What, if any, consistency is expected between national data and New York State?
    - What are the differences between New York State's testing program and the National Assessment of Educational Progress (NAEP)?
    - Should the percentage of students in each performance level be similar, even though the testing programs are not similar?
    - What differences between the results for New York State's testing program and the NAEP testing program are acceptable?

After the presentation of these guiding questions, panelists reviewed a description and impact data for a number of New York State examinations and for the NAEP. The panelists reviewed the results of the following assessments:

- Grade 8 Mathematics Test
    - Overall Test Level
    - Geometry Standard Level Based on three levels of the Student Performance Index (SPI)
- Regents Examination in Mathematics A
- Regents Examination in Mathematics B
- Regents Examination in Integrated Algebra
- National Assessment of Educational Progress (NAEP)
    - National Level Data, Grade 4 Mathematics Test
    - New York State Level Data, Grade 4 Mathematics Test
    - National Level Data, Grade 8 Mathematics Test
    - New York State Level Data, Grade 8 Mathematics Test

Panelists then broke into three groups of 10 panelists each. The groups met in separate rooms, where they first discussed reactions to the data just presented. They discussed each assessment in turn and how each is related to the Regents Examination in Geometry.

After this discussion, panelists were instructed to recommend the acceptable percentage of students who should be in each achievement level, using the rating sheet shown in Appendix B. Panelists were instructed to think about the design of the Regents Examination in Geometry, the design of the other assessments, the data that was presented, and the discussion of the guiding questions. Panelists made a “high” recommendation and a “low” recommendation. They completed the first round of recommendations before breaking for lunch.

After lunch, the committee reconvened to review the results from the first round. The committee was shown the minimum, maximum, mean, and median for each group and for the committee. A spokesperson from each group then summarized the discussion that had occurred in that group before the round one recommendations. Panelists were allowed to ask questions of other panelists and of NYSED staff members.

Then, panelists again broke into three groups and completed round two recommendations. For those panelists whose schedules permitted, the final average recommendations were presented without further discussion.

Results

The mean recommended percentage of students was computed by averaging “low” and “high” recommendations across all panelists. The median

was computed in a similar fashion. Table 1 summarizes, for round one, the panelists' recommendations for the acceptable percentage of students who should be classified as Level 2 and above and the acceptable percentage of students who should be classified as Level 3.

Table 1. Round 1 Results for the Recommended Percentage of Students Who Should Be Classified as Level 2 and Above or Level 3

		Level 2 and Above	Level 3
Group 1	Mean	74	21
	Median	75	20
	Standard Deviation	6	7
	Minimum	60	10
	Maximum	85	30
Group 2	Mean	84	20
	Median	85	20
	Standard Deviation	6	8
	Minimum	70	5
	Maximum	95	35
Group 3	Mean	79	23
	Median	80	23
	Standard Deviation	6	5
	Minimum	70	15
	Maximum	90	35
Committee	Mean	79	21
	Median	80	20
	Standard Deviation	7	7
	Minimum	60	5
	Maximum	95	35

Table 2 summarizes, for round two, the panelists' recommendations for the acceptable percentage of students who should be classified as Level 2 and above and the acceptable percentage of students who should be classified as Level 3.

Table 2. Round 2 Results for the Recommended Percentage of Students Who Should Be Classified as Level 2 and Above or Level 3

		Level 2 and Above	Level 3
Group 1	Mean	77	22
	Median	78	20
	Standard Deviation	5	5
	Minimum	65	15
	Maximum	85	30
Group 2	Mean	85	23
	Median	85	25
	Standard Deviation	6	7
	Minimum	70	5
	Maximum	95	35
Group 3	Mean	82	23
	Median	80	25
	Standard Deviation	6	6
	Minimum	70	10
	Maximum	95	35
Committee	Mean	81	23
	Median	80	25
	Standard Deviation	7	6
	Minimum	65	5
	Maximum	95	35

Figure 1 through Figure 4 present the impact data for each of the three achievement levels for both rounds, based on mean and median recommendations. Note that percentages may sum to more or less than 100% due to rounding values for presentation.

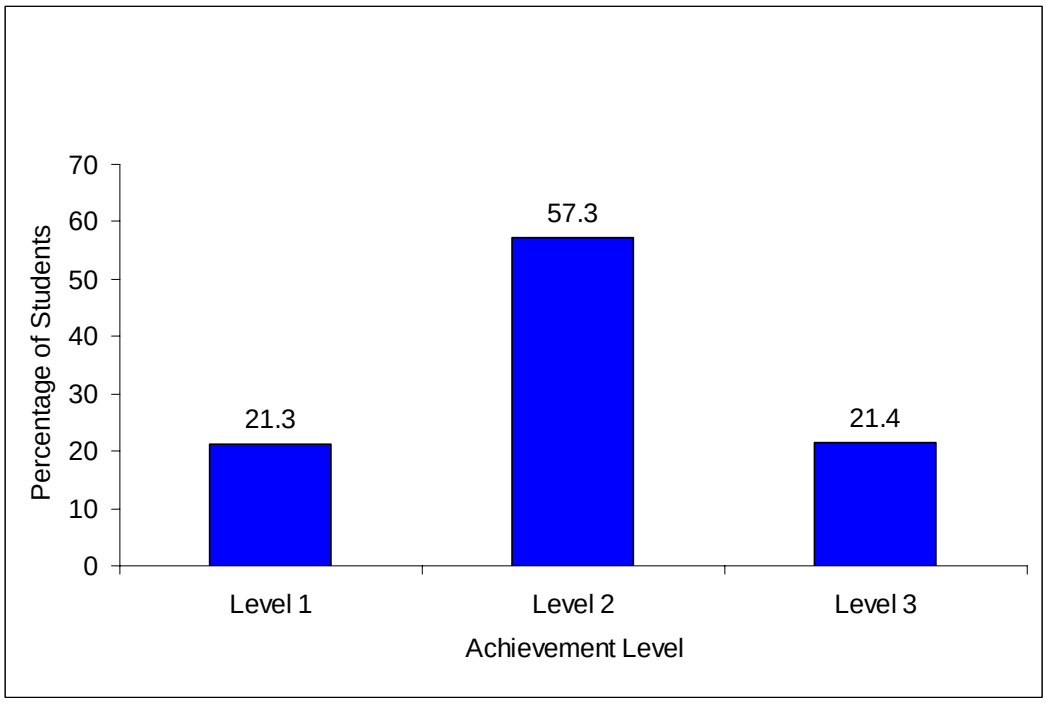

Figure 1. Mean Percentage of Students in Each Achievement Level After Round 1

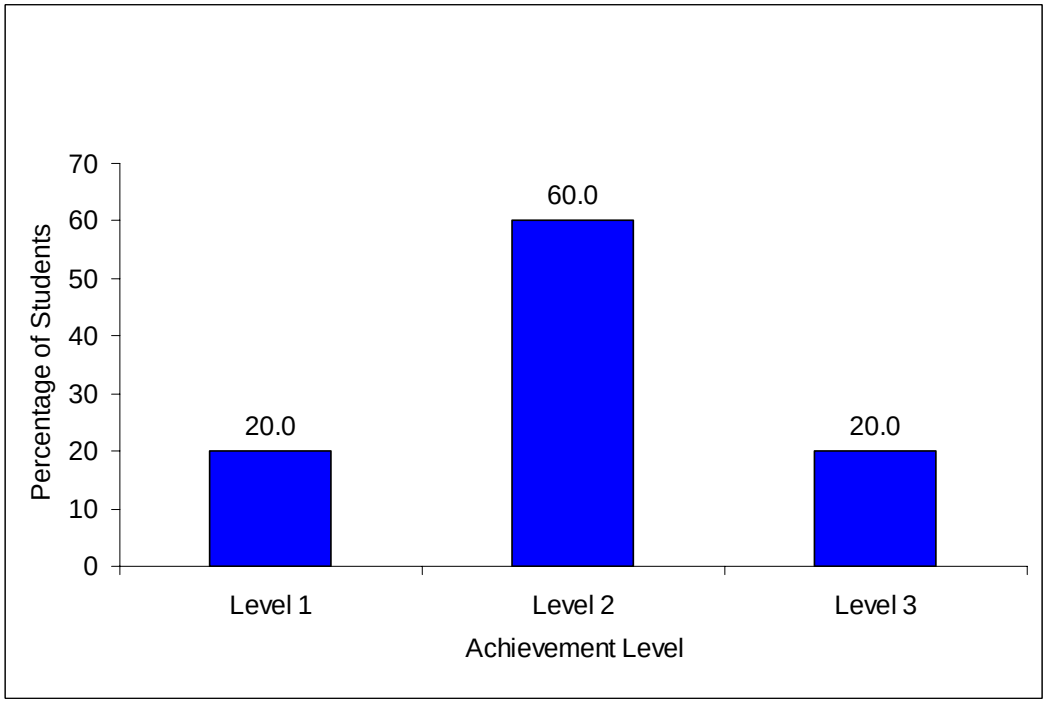

Figure 2. Median Percentage of Students in Each Achievement Level After Round 1

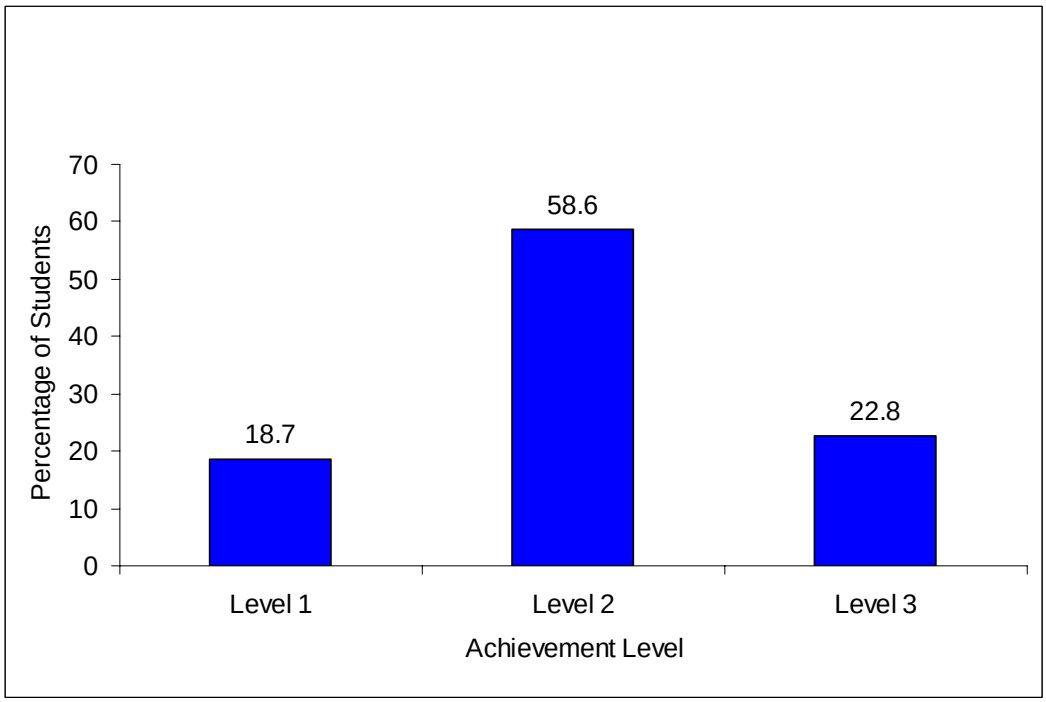

Figure 3. Mean Percentage of Students in Each Achievement Level After Round 2

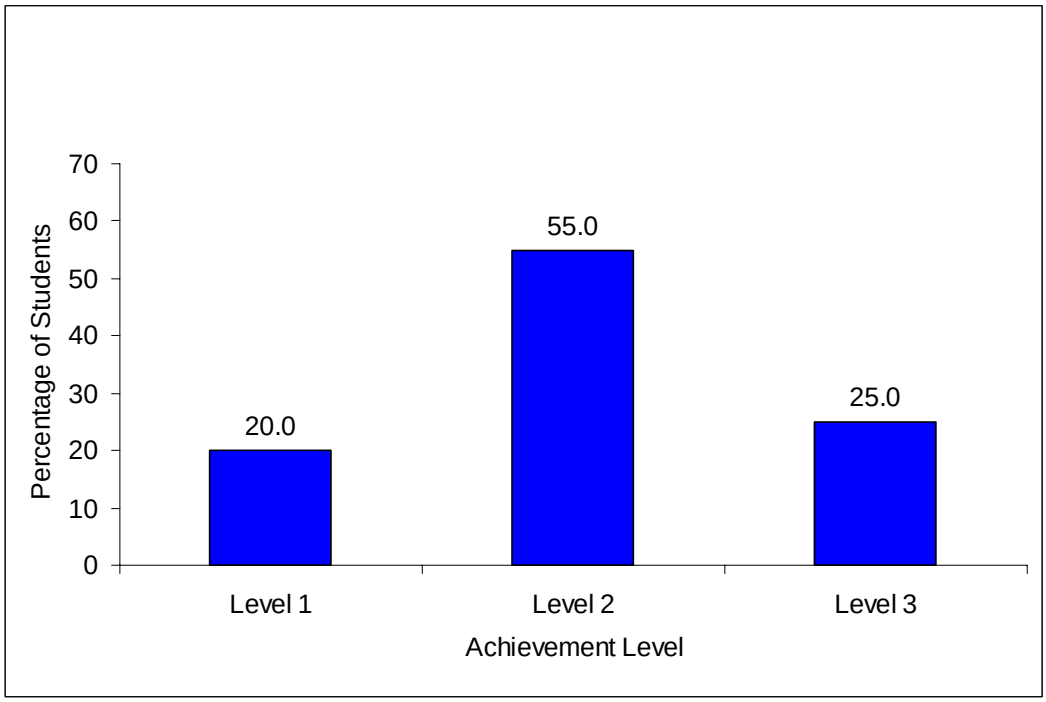

Figure 4. Median Percentage of Students in Each Achievement Level After Round 2

Evaluations

Exit surveys were completed by each panelist at the completion of the Pre-Policy Measurement Review Panel meeting. The exit survey was completed by 29 out of the 30 panelists. Panelists answered each question, using a scale of 1–5, 1 being "**totally disagree**" and 5 being "**totally agree**." The survey questions and the results are shown in Table 3.

Table 3. Questionnaire Results for the Pre-Policy Measurement Review Panel Meeting (N = 29)

Question	Mean	Median	Maximum	Minimum
1. The method for making recommendations on the ideal percent of students who should be classified in each achievement level was conceptually clear.	4.17	4.00	5	2
2. I had a good understanding of the design of the New York State Regents Examination in Geometry.	3.83	4.00	5	2
3. I had a good understanding of the design for the other assessments presented, such as NAEP.	3.31	3.00	5	1
4. After the first round of ratings, I felt comfortable with the method for making recommendations.	4.14	4.00	5	2
5. After the second round of ratings, I felt comfortable with the method for making recommendations.	4.34	4.00	5	3
6. I found the feedback on the recommendations of other panelists useful in making my own recommendations.	4.45	5.00	5	3
7. I found the feedback on the overall group recommendation useful in making my own recommendations.	4.34	4.00	5	3
8. I feel confident that the final cut score recommendations reflect the achievement levels associated with the Regents Examination in Geometry.	3.97	4.00	5	3

Item Mapping Standard Setting

Two committees of New York State educators convened June 22 and June 23, 2009, in Albany, New York, to recommend standards for the New York State Regents Examination in Geometry. The first committee, Committee A, had 29 educators, and the second committee, Committee B, also had 29 educators. The item mapping procedure was applied to recommend the cut scores.

Panelists

All panelists provided voluntary demographic information. Table 4 presents a summary of gender representation across both committees, Table 5 provides a summary of the ethnic representation of both committees, Table 6 lists the distribution of geographic locations of the panelists, and Table 7 summarizes the educational experience distribution between the two committees.

Table 4. Number of Male and Female Panelists in Committees A and B

	Committee A	Committee B
Female	18	20
Male	11	9

Table 5. Summary of the Ethnic Representation of the Panelists in Committees A and B

	Committee A	Committee B
Caucasian	19	16
Hispanic	2	3
African American	5	3
Other	3	2
Did not respond	0	5

Table 6. Distribution of Geographic Locations of Panelists for Standard Setting

	Committee A	Committee B
North Country	2	2
Long Island	2	0
New York City	4	10
Lower and Mid Hudson Valley	4	5
Capital Region	4	1
Central New York	5	5
Western New York	8	6

Table 7. Education Roles of Panelists in Committees A and B

	Committee A	Committee B
Mathematics Teacher	25	22
Special Education Teacher	1	1
Bilingual Teacher	0	2
Curriculum/Department/ Test Coordinator	2	4
Math Department Chair	1	0

Method

Panelists used an item mapping methodology, sometimes referred to as a bookmark approach, to recommend standards for the Regents Examination in Geometry. The item mapping methodology is typically conducted by using the following materials:

- Achievement level descriptors (ALDs)
- Ordered item books
- Item map

A description of each one is provided to give background for a description of the item mapping methodology. After the description of these materials, a description of the typical item mapping methodology is presented.

Achievement Level Descriptors

Standard setting panelists are tasked with estimating the performance of a group of students, e.g., the Basic, Proficient, or Advanced student. Students are grouped into these achievement levels as a way to establish and communicate achievement goals. The achievement levels define what students should know and be able to do when they have reached these achievement levels. For example, what should a student who has reached the Proficient level know and be able to do? States or other test developers create descriptions of what students should know and be able to do at different achievement levels. These descriptions are called achievement level descriptors (ALDs).

Generally, achievement levels represent a broad range of achievement. For example, more than one fourth of the students in a grade level for a state may be classified as failing within the Basic achievement level.

The general ALDs that attempt to capture the range of achievement represented by achievement levels are too vague for standard setting panelists tasked with estimating the performance of students in each achievement level. Panelists make ratings of items, student work samples, or students using descriptions of what students know and can do at each achievement level. Panelists need descriptions that contain enough detail to support reliable ratings between panelists, across occasions, and across panelists.

To support reliable ratings in standard setting, descriptions of what students just into the Proficient achievement level or just into the Advanced achievement level know and can do are created. These students that are just Proficient or just Advanced are known as threshold examinees because they define the threshold of the achievement level. Threshold examinees are students with the minimum level of proficiency needed to make it into a particular achievement level.

The descriptions of what just Proficient or just Advanced students know and can do play a central role in standard setting. The panelists are instructed to use these ALDs of what just Proficient or just Advanced students know and can do as the frame of reference for each judgment. The construct being measured is the panelists' representation of just Proficient or just Advanced students' performance. The measurement of that construct results in cut points recommended by panelists.

The logic of using ALDs for threshold students to delimit the range of achievement represented by achievement levels is straightforward. The ALDs for threshold students describe what the most minimally qualified student in that achievement level knows and can do. Students who are not likely to know or be able to do what the threshold students know and can do must fall into the previous achievement level. Students who are likely to know or be able to do

more than what the threshold students know and can do must fall into the current or succeeding achievement levels.

Ordered Item Book

Under the item mapping method, panelists review test items from least to most difficult. Panelists are typically given a book of test items, called an ordered item book, to help them with this review. The items in this book are presented one item per page and are ordered from the least difficult items to the most difficult. Often, a three-ring binder is used for the ordered item book.

The ordered item book may include selected-response items and constructed-response items. Each selected-response item, such as a true/false or multiple-choice item, is presented only once in the book. A multiple-choice item page will show the test item stem and alternatives, as well as the correct response. A true/false item page will show the test item and the correct response.

Each constructed-response item is presented multiple times, corresponding to the number of score points in the rubric. Each score point for a constructed-response item is presented once in the book, except the 0 score point. For example, a constructed-response item that is scored using a 4–point rubric (0–4) would have four pages in the ordered item book representing score points 1, 2, 3, and 4. The page for each score point or item step will present the prompt and an example of student work awarded that particular score point. The example of student work should be a clear representation of performance at that score value. The rubric used to score student performance should also be available.

For example, an ordered item book might be constructed for an assessment with 30 multiple-choice items and 8 constructed-response items, each scored on a scale of 1–3. The ordered item book would include 30 pages, 1 page for each of the 30 multiple-choice items. In addition, the ordered item book would include 24 pages, 1 page for each of the three score points for each of the 8 constructed-response items. The ordered item book would total 54 pages.

Sometimes an ordered item book is constructed by using more items than the number of items on an assessment. The items in an ordered item book should represent the categories of content, mix of item formats, and range of difficulty described in the test blueprint. Items from the item bank may be added to provide a better representation of the test blueprint. For example, items from a content category might be added if that category was not fully represented on a test form. Alternatively, items from the item bank may be added so that items represent the entire scale range. For example, the ordered item book may have a sequence of items with difficulty values of 0.00, 0.50, and 1.00 logits. Items

with difficulty values near 0.25 and 0.75 logits may be added to the ordered item book to represent the gaps in the scale between items on the test form.

The empirical order of item difficulty must be calculated before the ordered item book can be constructed. Empirical difficulty represents a point on a known ability scale. The ability scale is commonly established by using Item Response Theory under a Rasch or combined model.

Empirical difficulty is calculated for selected-response and constructed-response items. Selected-response items include true/false items and multiple-choice items. The empirical difficulty for selected-response items is calculated as the point on the ability scale at which the examinee would have a given probability, called a response probability (RP), of selecting the correct response. Guessing should be factored out of the response probability when computing the empirical difficulty.

Empirical difficulties are computed for those constructed-response items that are scored using a rubric. Constructed-response items are represented by multiple score points, corresponding to the number of score points in the rubric. The empirical difficulty for each score point is calculated as the point on the ability scale at which the examinee would have a given RP of achieving at least that score point. This definition of empirical difficulty for constructed-response score points is conceptually similar to the definition of empirical difficulty for selected-response items. Note that the empirical difficulty should be greater for higher score points than for lower score points. A score point of at least 3 will be more difficult to obtain than a score point of at least 2.

The Regents Examination in Geometry contains twenty-eight selected-response items (multiple-choice items) and ten constructed-response items. The selected-response items are weighted by 2 for scoring, and the constructed-response items are weighted by 1. For the ten constructed-response items, six items have a score range from 0 to 2, three items have a score range from 0 to 4, and one item has a score range from 0 to 6. The raw scores for the Regents Examination in Geometry range from 0 to 86.

Rasch and Partial Credit Models

The Rasch model and the Partial Credit model are used for all the Regents examinations. The Rasch model is applied to fit the multiple-choice items, and the Partial Credit model is applied to fit constructed-response items. Research in standard setting methodology tends to indicate that when an RP value of 0.67 is used, the maximum information needed for standard setting can be achieved. In addition, the RP value of 0.67 has been used historically for other assessments in New York State, such as the Grades 3–8 assessments. This value was also applied when empirical item difficulty of the items was calculated to construct the ordered item book. The Rasch model and the

computation of empirical difficulty value with an RP value of 0.67 are discussed below.

When it is a dichotomous item, the Rasch model can be defined as the following:

$$P = \frac{1}{1+e^{-(\theta-b)}}$$

Using the operational data, item difficulty parameter b was calibrated using WINSTEPS. Based on the theory of the Rasch model, the item difficulty parameter b from the calibration corresponds to a proficiency θ value when the RP value is 0.50. To obtain the item parameter value and hence the corresponding θ value that will have an RP value of 0.67, modification needed to be conducted on the item parameters. Basically, the following equations needed to be solved for b', the item difficulty, hence the ability level for an RP value of 0.67.

$$0.50 = \frac{1}{1+e^{-(\theta-b)}}$$

$$0.67 = \frac{1}{1+e^{-(\theta-b')}}$$

Solving this equation results in a value for b', where $b' = b + \ln 2 = b + 0.69315$. Therefore, a factor of 0.69315 was added to the multiple-choice item parameters (dichotomous items only) for the items to be included in the ordered item book.

When it is a polytomously scored item (constructed-response item), the formulas are a bit more complicated. The IRT Partial Credit model was used to analyze polytomously scored constructed-response items for the New York State Regents Examinations. The model is defined as

$$P_{xi} = \frac{\exp\sum_{j=0}^{x}(\theta - D_{ij})}{\sum_{k=0}^{m_i}\left[\exp\sum_{j=0}^{k}(\theta - D_{ij})\right]},$$

where $x = 0,1,...,m_i$. D_{ij} values were available from the calibration of operational data, and they were obtained using a response probability of 0.50, by model definition.

To obtain RP = 0.67 difficulty values, more intensive computation needed to be conducted to produce the value. It was more complicated than a simple addition factor, as is the case with dichotomously scored items. The idea was to produce the ability value that would yield a probability of 0.67 for a given score category and above. Basically, the ability value associated with a score value of 2 for a 4-point item indicates the ability that will yield a probability of 0.67 for a

student to get a score of at least 2 (including 2, 3, and 4) for this 4-point item. To conduct this computation, an iterative process was employed, with θ in the increment of 0.001, to locate the corresponding b' value that would yield the RP value of 0.67. The b' value was computed for all score points for each of the constructed-response items. Two independent psychometricians conducted the analysis, and their results were a 100% match.

After all the values were computed, the ordered book was created by ordering the items in terms of the computed b' values. In addition, items from two anchor forms were included in the ordered item book to include more content and statistical coverage for the test. The ordered item book is located in Appendix I. There were 90 pages in the ordered item book. Table 8 summarizes the composition of the ordered item book.

Table 8. Composition of the Ordered Item Book

	Number of Items	Maximum Credit	Number of Pages
Operational Test			
Multiple Choice	28	1	28
2-credit Item	6	2	12
4-credit Item	3	4	12
6-credit Item	1	6	6
Anchor Forms			
Multiple Choice	14	1	14
2-credit Item	4	2	8
4-credit Item	1	4	4
6-credit Item	1	6	6

Item Map

The item map is a handout that accompanies the ordered item book and provides additional information for each item. The item map is a table that consists of one row for each item in the ordered item book. The items are listed on the item map in the same order they are presented in the ordered item book; i.e., from least to most difficult based on the empirical item difficulty calculated using an RP value of 0.67. Each row lists information about the item. The following information is commonly provided for each item:

- The page number in the ordered item book
- The original item number on the test form (unless the item is from the test bank)
- The content classification of the item
- The key (unless the row corresponds to a score point for a constructed-response item)
- The maximum score point (if the item is a constructed-response item)

After round one of the standard setting procedure, an augmented item map is often distributed to panelists as part of the structured feedback provided between rounds of ratings. The augmented item map presents the information from the original item map and adds information about item difficulty, typically the percentage of students who answered the item correctly (for multiple-choice items) or the percentage of students who earned each score point or higher (for constructed-response items).

Item Mapping Methodology

Under the item mapping standard setting method, panelists are asked to review items in the ordered item book and make a judgment as to the likelihood of threshold examinees answering an item correctly or achieving a given score point or higher. This judgment is made within a given frame of reference, for a given RP value, and within a given procedure.

The panelists are instructed to use the ALDs as the frame of reference for each judgment. The panelists have completed a warm-up task to become familiar with the ALDs. Sometimes, the panelists may have created the ALDs during an earlier session. These ALDs describe what the threshold examinees at each achievement level (e.g., just Level 2 or Level 3) know and can do. Panelists use only one ALD at a time.

Panelists are instructed to judge the likelihood of threshold examinees answering an item correctly or achieving at a given score point. The RP value used for this assessment was 0.67. Panelists may be instructed to think of this RP value in several ways. Panelists may be instructed to think about a group of 100 threshold students (e.g., just Proficient students). For an RP value of 0.67, panelists are asked to identify the item that 67 of 100 threshold students will answer correctly. Alternatively, panelists may be instructed to think of a typical threshold student, perhaps a student they are teaching or have taught. Again for an RP value of 0.67, panelists are asked to identify the item that this student would have a 67% chance of answering correctly.

The task set for panelists is to read each item or score point in the ordered item book and evaluate the knowledge, skills, and abilities required to respond correctly to the item or to produce a response at the score point. Panelists then compare their evaluation of the cognitive demands of each item and score point to the assigned ALD, e.g., the description of the just Proficient examinees. Panelists should proceed from the least difficult items to the most difficult. Keeping in mind the ALD, panelists are instructed to identify the last item or score point that 67 of 100 threshold students should answer correctly. For the item immediately following, panelists should judge that only 66 or fewer of 100 just Proficient examinees would respond correctly. For the item immediately preceding, panelists should judge that 68 or more of 100 just Proficient examinees would respond correctly. Panelists then mark that page in the

ordered item book, the last yes page, often using a self-adhesive note, and record the item identifier on a record sheet.

Cut Score Computation

The cut score at each achievement level was determined by computing the median from the judge ratings. For a given achievement level, each judge for each round had a page number recommendation. These page numbers then were translated into Rasch values where an ability of this level produces an RP value of 0.67 of answering the item correctly. The median of these Rasch values was then computed, which was the cut score recommendation on the θ scale. The raw to θ conversion table was used to look up the corresponding raw cut score. The standard setting θ was likely to be between two θ values on the raw to θ conversion table. To give students the benefit of the doubt, based on NYSED's direction, the lower of the two θ was identified, and its associated raw score was used for the raw cut score recommendation.

This identified raw score represents the minimum raw score that an examinee must attain to be classified into a particular achievement level based on the standard setting methodology. As mentioned before, the ordered item book contained 90 pages representing 90 score points—including operational items and items from the two anchor forms. The raw to θ conversion table was based on the operational Regents Examination in Geometry and had raw scores ranging from 0 to 86. The median panelist rating was computed for each achievement level. Using that median ability value, the corresponding raw score was identified.

For example, at round three, the median page number for Level 2 in group A was 30. The item on page 30 had the θ value of 0.666 (see Appendix I, ordered item book). Next, in the raw to θ conversion table, a raw score of 50 corresponds to a θ value of 0.664, and a raw score of 51 corresponds to a θ value of 0.707. Per NYSED's direction, the raw cut score recommended then was 50. The rest of the cut scores were identified using the same algorithm.

Methodological Strengths

The item mapping method has several features that make it an appealing standard setting approach. First, the item mapping method can be used with a mixed-format assessment. Panelists consider selected-response items and constructed-response items when placing bookmarks. Consequently, panelists' cut score recommendations reflect the mix of item formats found on a test.

Second, the task that panelists complete within the item mapping method may be relatively less challenging than the panelists' task under other standard setting methods. Proponents of the item mapping method argue that panelists are required to make relatively few judgments compared with the number of

judgments required of panelists under other standard setting methods. For example, panelists using the item mapping method to recommend cut scores for three achievement levels would be required to make only two judgments. In contrast, panelists using an Angoff method to recommend cut scores would be required to make one judgment for each item.

In addition, panelists using the item mapping method are required to spend relatively less time reviewing the test items. A panelist who has reviewed the first group of items and placed the first bookmark need not review those items again to place a subsequent bookmark. The panelist would place the first bookmark and then continue paging through the ordered item book to find the appropriate item on which to place the next bookmark.

Before an item mapping procedure can be conducted, substantial work must be done, including collecting student responses and calibrating and scaling items, using Item Response Theory. Student responses may be collected through either a field test or an operational administration. An operational administration is more likely to provide a larger number of responses, collected under more realistic conditions, than a field test.

Procedure

The Recommendations for Setting Achievement Levels for the Regents Examination in Geometry meeting began on Monday, June 22. The agenda for the meeting is shown in Appendix D. The morning was devoted to staff introductions, a description of standard setting, and a description of the Regents Examination in Geometry. Altogether, 58 educators participated in the conference.

After the midmorning break, all the educators remained in the same room and began the process of reviewing ALDs. This activity (the two independent committees discussing the ALDs together in one large group before the standard setting process) was recommended by the Technical Advisory Committee (TAC). The purpose of the activity was to make sure both committees were using the same expectations for students in each of the achievement levels when recommending achievement standards. This process required several hours and resulted in a set of descriptors for each achievement level (Level 1, Level 2, and Level 3). Appendix E presents the general ALDs provided by NYSED. The educators then broke into eight small groups and discussed specific ALDs. After the small group discussions, all panelists reconvened, and each group presented its ALDs. A typed summary of the ALDs was captured from the discussions and made available to each of the educators for the rest of the standard setting conference. Appendix E also provides the specific ALDs by the educators. The specific ALDs consisted of two parts: descriptors for each of the three achievement levels, and the most distinguishing features for the students who are at the threshold of each achievement level, Level 2 and Level 3.

After the discussions about ALDs and after the educators agreed on the general expectations of what students should know and be able to do in each of the achievement levels, the large group was broken into two separate committees. From here on, the standard setting process was independent between the two committees.

Each committee met in its own meeting room and began the standard setting process. There were twenty-nine educators per committee, and these educators were pre-assigned to four different tables. A leader was assigned for each table. The item mapping procedure was the methodology used. Panelists were instructed to identify the last item in an ordered item book that a threshold student at a given level would have a response probability of at least 0.67 of answering correctly.

The ordered item books were constructed from operational items from the June 2009 administration and anchor items from the 2008 field test administration. Items were sorted from least to most difficult, using the Rasch item difficulty values based on an RP value of 0.67.

The standard setting process consisted of three rounds of judgments. The recommendation form used by the panelists is shown in Appendix F.

Panelists were provided with feedback between each round. The feedback was intended to inform the panelists' decisions, but not to dictate their ratings. After round one, panelists met in small groups of seven or eight people. They were provided the cut scores (in terms of ordered item book page number) for each panelist on the basis of the round one ratings in addition to the mean, median, minimum, and maximum cut score at each level for that table. In reviewing the cut score report, individual panelists were asked to think about the following:

- How similar are your cut scores to that of the group? (i.e., Is a given panelist more lenient or stringent than the other panelists?)
- If so, why is this the case? Do panelists have different conceptualizations of these borderline students? Were ALDs being used when making the ratings?

Panelists were informed that there was no intention for them to come to a consensus on cut score judgments, but they should discuss differences to gain an understanding for why differences exist.

In addition, panelists were provided a list of item *p*-values. Finally, panelists were presented with the raw cut score based on their committee's round one rating. The *p*-values were based on a representative sample of approximately 124,838 students who took the operational exam in June 2009.

Within each committee, panelists were given time to discuss the appropriateness of the committee-level cut scores, given the proportion of students that would fall into each level.

After round two, panelists received the cut scores for each panelist on the basis of the round two ratings, in addition to the mean, median, minimum, and maximum cut scores at each level for that table. Next, panelists were given the mean, median, minimum, and maximum cut scores for the committee (across tables). The facilitator led the discussion with all four tables combined. The facilitator noted the similarities and differences across tables but reminded the panelists that a consensus was not required.

Next, panelists were provided with the overall cut score on the raw score metric, as well as a graphical display of the percentage of students in each achievement level on the basis of the median cut scores from round two. The impact data was based on the same representative sample on which the *p*-values were based. Panelists were also provided with a graphical display of the percentage of students in each achievement level disaggregated for Grade 9 students and Grade 10 and above students.

Within each committee, the panelists were given time to discuss the appropriateness of the committee-level cut scores, given the proportion of students that would fall into each level.

After the panelists had a chance to discuss their current cut score recommendations and the related impact, they provided the rating for round three, the final round. The median from round three ratings from each committee was considered the final cut score recommendation for the committee.

After round three rating and analysis, both committees reconvened. The final round recommendations from both committees, along with their impact, were presented. Next, the panelists were instructed to fine-tune the ALDs they had developed on the first day, before the standard setting activity. The edits and the final ALDs were captured. They are provided in Appendix E.

After completion of the editing of ALDs, the panelists filled out exit surveys, were thanked for their time and participation, and were dismissed. The table leaders from each committee, a total of eight people, were asked to stay and participate in the synthesis meeting. The synthesis meeting was scheduled based on the advice from TAC, and the purpose of the synthesis was to focus on the differences in the cut score recommendations from the two independent committees. In fact, after round three, the two committees provided exactly the same cut score recommendations, as presented in the following results section. Still, the synthesis group met and focused on the items that were around the cut score recommendations. These eight panelists focused on the knowledge and skills those items were measuring, and how they related to the ALDs and

especially the differences between the students who were just below the achievement level and the students who were just above the achievement level (the threshold students). The synthesis group members then made their final recommendation.

Results

Table 9 summarizes cut score recommendations in terms of page number, as well as raw cut scores for achievement Level 2 and Level 3 for round one. Table 10 summarizes cut score recommendations for round two, and Table 11 presents the final cut score recommendations for round three. For each round, the mean, median, minimum, and maximum page number recommendations are presented, as well as the raw cut score recommendation based on the median recommendation from the entire committee. As can be observed from the tables for each round, the cut score recommendations on the raw score metric were very consistent between the two independent committees. Discussions on ALDs with the two committees combined probably contributed to the consistency between the two committees.

Comparisons across rounds also indicate that the cut score recommendations did not fluctuate much between rounds—basically around 1 or 2 points on the raw score metric. Item empirical difficulty (*p* values) was presented after round one, and impact data (the percentage of students in each achievement level based on the cut score recommendation) was presented after round two. These two pieces of additional information seemed to have no great effect on the overall cut score recommendations in either of the two committees. Standard deviations are not presented in these tables because, as the previous section indicated, all the computations were conducted at the θ metric and translated back to either page number or raw scores. With mean, median, minimum, and maximum values, the translation worked well; but with standard deviation, the translation would not have worked well. Therefore, standard deviations were not provided.

Table 9. Cut Score Recommendations by Committee for Level 2 (65–84) and Level 3 (85–100), Round 1

		Page Number				Raw
		Mean	Median	Minimum	Maximum	
Committee A	Level 2	30	33	18	47	50
	Level 3	66	68	36	82	71
Committee B	Level 2	37	37	12	66	51
	Level 3	71	70	21	84	72

Table 10. Cut Score Recommendations by Committee for Level 2 (65–84) and Level 3 (85–100), Round 2

		Page Number				Raw
		Mean	Median	Minimum	Maximum	
Committee A	Level 2	27	32	17	43	50
	Level 3	66	66	41	80	69
Committee B	Level 2	36	32	19	49	50
	Level 3	69	69	39	79	71

Table 11. Cut Score Recommendations by Committee for Level 2 (65–84) and Level 3 (85–100), Round 3

		Page Number				Raw
		Mean	Median	Minimum	Maximum	
Committee A	Level 2	27	30	17	43	50
	Level 3	66	67	45	80	71
Committee B	Level 2	27	30	19	49	50
	Level 3	68	69	42	79	71

Figure 5 and Figure 6 present the percentage of students in each achievement level using the cut score recommendations after rounds two and three. The impact data were based on a representative sample of 124,838 students who participated in the operational testing of the June 2009 Regents Examination in Geometry administration. These figures were presented to the panelists after round two and round three, respectively. In order to keep the two committees totally independent during the standard setting process, each

committee was only presented the impact data based on its own recommendation.

Not surprisingly, the two committees had exactly the same impact data based on the round three rating because their cut score recommendations were identical.

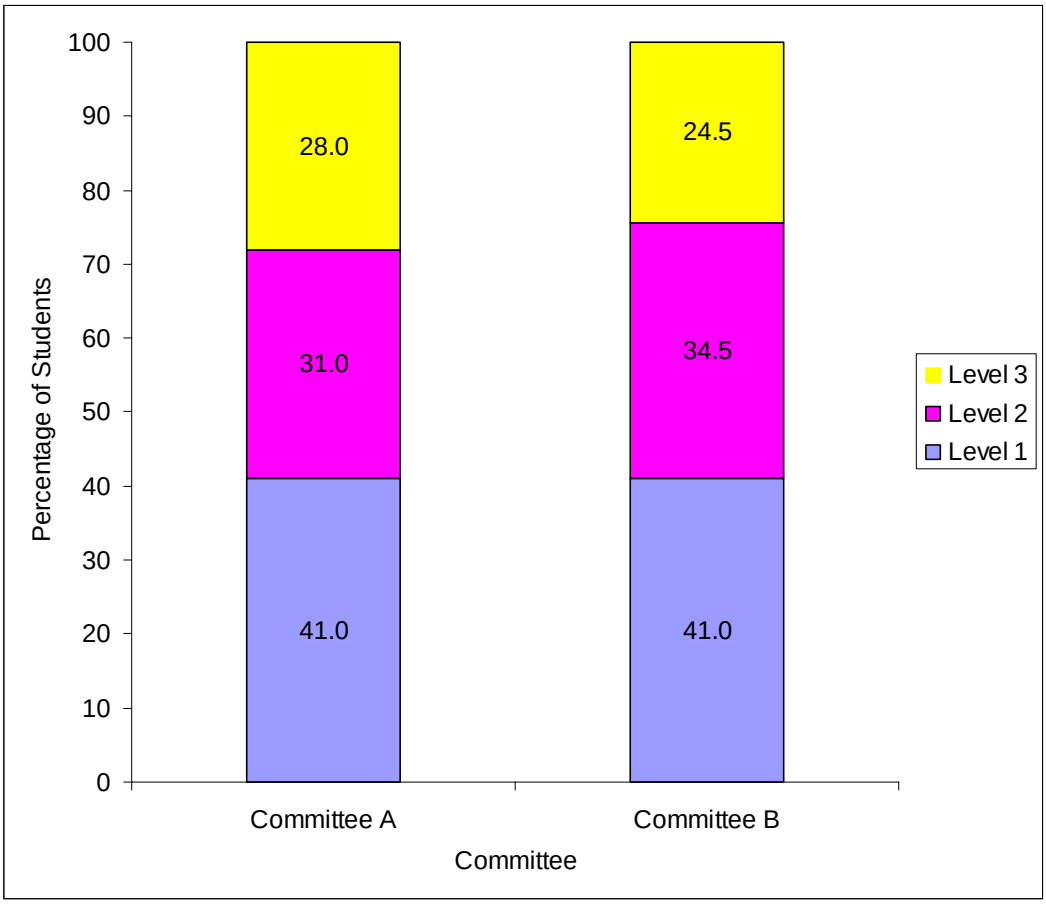

Figure 5. Percentage of Students in Each Achievement Level Based on Each Committee's Cut Score Recommendations After Round 2

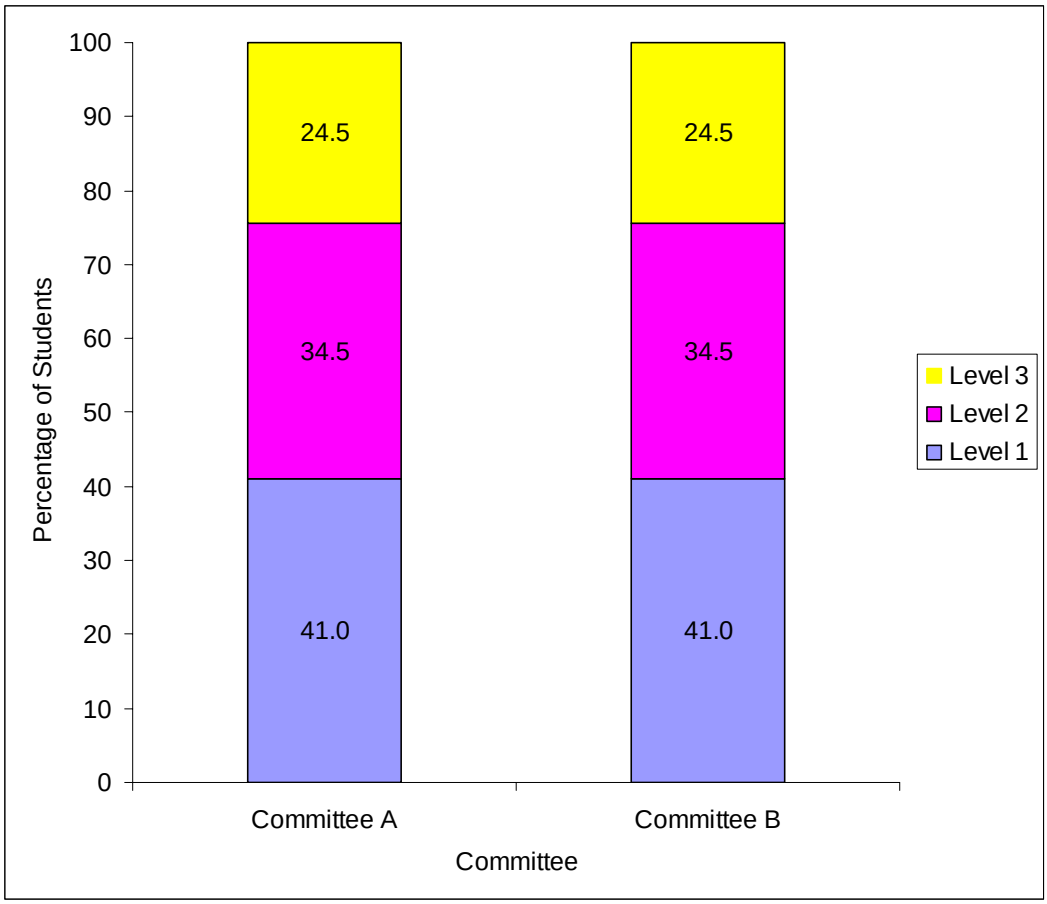

Figure 6. Percentage of Students in Each Achievement Level Based on Each Committee's Cut Score Recommendations After Round 3

After the two independent standard setting committees A and B completed their final recommendations and finalized the ALDs, a synthesis group was convened. The table leaders were invited to stay to participate in the synthesis group. There were eight table leaders who participated in this activity. The purpose of the synthesis was to further examine the differences (if there were any) between the cut score recommendations of the two independent committees, to discuss these differences, and to come to a final recommendation.

One panelist from each committee gave a brief description of the recommendations and the rationale behind the recommendations. Next, the panelists were asked to observe the cut score recommendations from both committees and to use the ordered item book to further look at the items identified as the bookmarks. The panelists were asked to observe the knowledge and skills the items at the cut and around the cut were measuring, to observe the related ALDs for borderline "just make it" students for each achievement level, and to make an overall recommendation.

Only one round of rating was conducted for the synthesis meeting, with each panelist participating in the synthesis providing a rating for each of the two cuts: Level 2 (65–84) and Level 3 (85–100). Table 12 summarizes the results from the synthesis meeting.

Table 12. Cut Score Recommendation from the Synthesis Meeting

	Page Number				Raw
	Mean	**Median**	**Minimum**	**Maximum**	
Level 2	49	50	46	51	50
Level 3	71	71	65	71	71

Panelist Variability

In order to describe the variability in panelists' judgments, a Generalizability Theory study was performed. This information could be used to determine how similar the cut scores might be if a different set of panelists or different composition of small groups was used to set cut scores. For this investigation, the sources of variability of interest were panelists, small groups, and rounds. For each cut score, the variance associated with each of these sources was estimated using the maximum likelihood SAS VARCOMP procedure. For this study, the number of rounds was treated as a fixed factor (3 rounds in total, a typical practice in standard setting meetings), meaning that if the standard setting meeting was held again, the same number of rounds would be used. In addition, because judges discussed all activities in small groups, their judgments were considered dependent on group membership. Therefore, judges were considered "nested" within tables.

The judge variability estimates based on Generalizability Theory are presented in Table 13 through Table 16.

Table 13. Generalizability Theory Analysis of Judge Variability, Level 2 Cut Score, Committee A

Variance Component	Estimated Variance Component	Applied Variance Component	Percent of Variance	Error Variance	Standard Error
Table	1.25630	1.25630	9		
Judge:Table	5.97706	5.97706	41		
Round	0.38368	0.38368	3		
Table x Round	0.34848	0.34848	2		
Remaining	6.58968	6.58968	45		
				0.38454	0.62011

Table 14. Generalizability Theory Analysis of Judge Variability, Level 3 Cut Score, Committee A

Variance Component	Estimated Variance Component	Applied Variance Component	Percent of Variance	Error Variance	Standard Error
Table	26.2199	26.2199	45		
Judge:Table	19.7860	19.7860	34		
Round	-0.2195	0.0000	0		
Table x Round	0.3692	0.3692	1		
Remaining	12.1492	12.1492	21		
				6.76045	2.60009

Table 15. Generalizability Theory Analysis of Judge Variability, Level 2 Cut Score, Committee B

Variance Component	Estimated Variance Component	Applied Variance Component	Percent of Variance	Error Variance	Standard Error
Table	5.0764	5.0764	15		
Judge:Table	16.3473	16.3473	49		
Round	1.9593	1.9593	6		
Table x Round	-0.7189	0.0000	0		
Remaining	10.0240	10.0240	30		
				1.43883	1.19951

Table 16. Generalizability Theory Analysis of Judge Variability, Level 3 Cut Score, Committee B

Variance Component	Estimated Variance Component	Applied Variance Component	Percent of Variance	Error Variance	Standard Error
Table	16.3165	16.3165	31		
Judge:Table	19.9256	19.9256	37		
Round	-1.1581	0.0000	0		
Table x Round	3.2827	3.2827	6		
Remaining	13.7490	13.7490	26		
				4.29041	2.07133

Evaluations

An exit survey was completed by each panelist after the completion of standard setting. Panelists answered each question, using a scale of 1–5, 1 being "totally disagree" and 5 being "totally agree." The survey questions and the results for Committee A are shown in Table 17 and for Committee B in Table 18.

Table 17. Questionnaire Results for Committee A

Question	Mean	Median	Minimum	Maximum	Standard Deviation
1. The method for providing the rating was conceptually clear.	4.3	4.0	2.0	5.0	0.7
2. I had a good understanding of what the test was intended to measure.	4.2	4.0	2.0	5.0	0.8
3. I could clearly distinguish between student achievement levels.	4.4	4.0	3.0	5.0	0.6
4. After the first round of ratings, I felt comfortable with the standard setting procedure.	4.0	4.0	2.0	5.0	1.1
5. I found the feedback on *p*-values useful.	4.4	5.0	2.0	5.0	0.9
6. I found the feedback reports on the rating of panelists useful.	4.4	5.0	2.0	5.0	0.8
7. I found the feedback on the percentage of the students tested that would be classified at each achievement level useful.	4.3	4.0	3.0	5.0	0.7
8. Table discussion was open and honest.	4.8	5.0	3.0	5.0	0.5
9. I believe that my opinions were considered and valued by my group.	4.7	5.0	4.0	5.0	0.5
10. I am confident that my round 3 ratings for 65–84 and 85–100 reflect the knowledge, skills, and abilities described in the achievement level descriptors.	4.4	5.0	2.0	5.0	0.9
11. I am confident that the final cut score recommendations reflect the achievement levels associated with the New York State Regents Examination in Geometry.	4.1	4.0	2.0	5.0	1.0
12. I would defend the standards recommended by our committee.	4.3	5.0	2.0	5.0	1.0

Table 18. Questionnaire Results for Committee B

Question	Mean	Median	Minimum	Maximum	Standard Deviation
1. The method for providing the rating was conceptually clear.	4.4	5.0	2.0	5.0	0.7
2. I had a good understanding of what the test was intended to measure.	4.5	5.0	2.0	5.0	0.7
3. I could clearly distinguish between student achievement levels.	4.3	5.0	2.0	5.0	0.9
4. After the first round of ratings, I felt comfortable with the standard setting procedure.	4.4	5.0	2.0	5.0	0.7
5. I found the feedback on *p*-values useful.	3.8	4.0	1.0	5.0	0.9
6. I found the feedback reports on the rating of panelists useful.	4.3	4.0	3.0	5.0	0.6
7. I found the feedback on the percentage of the students tested that would be classified at each achievement level useful.	4.3	4.0	1.0	5.0	0.9
8. Table discussion was open and honest.	4.8	5.0	3.0	5.0	0.5
9. I believe that my opinions were considered and valued by my group.	4.7	5.0	3.0	5.0	0.6
10. I am confident that my round 3 ratings for 65–84 and 85–100 reflect the knowledge, skills, and abilities described in the achievement level descriptors.	4.6	5.0	3.0	5.0	0.6
11. I am confident that the final cut score recommendations reflect the achievement levels associated with the New York State Regents Examination in Geometry.	4.1	4.0	2.0	5.0	1.1
12. I would defend the standards recommended by our committee.	4.3	5.0	2.0	5.0	0.9

A decision factor survey was also completed by each panelist after the completion of standard setting. Panelists answered each question using a scale of 1–5, 1 being "**not at all**" and 5 being "**very strongly**." The decision factor survey questions and the results for Committee A are shown in Table 19 and for Committee B in Table 20. As can be observed from the tables, generally speaking, the most influential factors in panelists' decision making during the recommendations at standard setting appear to be their experience in education and their understanding of the ALDs.

Table 19. Decision Factor Survey Results for Committee A

Decision Factors	Mean	Median	Minimum	Maximum	Standard Deviation
1. Your experience in education	4.3	5.0	2.0	5.0	0.9
2. Before this item mapping standard setting, your perceptions about students in each of the three achievement levels	3.0	3.0	1.0	4.0	0.8
3. Your prior knowledge about standard setting	2.1	2.0	1.0	5.0	1.3
4. The orientation on standard setting presented today	3.7	4.0	1.0	5.0	1.1
5. Your perception of the high stakes versus low stakes context of the New York State Regents Examination in Geometry	2.8	3.0	1.0	4.0	1.1
6. Your thinking about students in each achievement level with whom you have had experience	3.7	4.0	2.0	5.0	0.9
7. The consequences of your decisions for No Child Left Behind Act	2.3	2.0	1.0	5.0	1.4
8. Your concerns about district, state, political, or economic issues	3.0	3.0	1.0	5.0	1.4
9. Your understanding of the achievement level descriptors	4.0	4.0	2.0	5.0	0.8
10. The item *p*-values that were presented after round 1	3.7	4.0	2.0	5.0	0.9
11. The impact data presented after rounds 1 and 2	3.8	4.0	2.0	5.0	0.9
12. The feedback report on estimated raw score cuts from rounds 1 and 2	3.7	4.0	2.0	5.0	1.0
13. Your interactions with your fellow panelists in your group before **round 1**	3.3	4.0	1.0	5.0	1.1
14. Your interactions with your fellow panelists in your group before **round 2**	3.6	4.0	1.0	5.0	1.0
15. Your interactions with your fellow panelists in your group before **round 3**	3.4	3.0	1.0	5.0	1.0
16. Your interactions with your fellow panelists in the large group discussion	3.3	3.0	1.0	5.0	1.0

Table 20. Decision Factor Survey Results for Committee B

Decision Factors	Mean	Median	Minimum	Maximum	Standard Deviation
1. Your experience in education	4.4	5.0	3.0	5.0	0.6
2. Before this item mapping standard setting, your perceptions about students in each of the three achievement levels	3.4	4.0	1.0	5.0	1.2
3. Your prior knowledge about standard setting	2.1	1.0	1.0	5.0	1.5
4. The orientation on standard setting presented today	3.5	4.0	1.0	5.0	0.9
5. Your perception of the high stakes versus low stakes context of the New York State Regents Examination in Geometry	3.4	3.0	1.0	5.0	1.1
6. Your thinking about students in each achievement level with whom you have had experience	3.9	4.0	2.0	5.0	1.0
7. The consequences of your decisions for No Child Left Behind Act	2.5	3.0	1.0	5.0	1.2
8. Your concerns about district, state, political, or economic issues	2.5	3.0	1.0	5.0	1.3
9. Your understanding of the achievement level descriptors	4.0	4.0	3.0	5.0	0.6
10. The item *p*-values that were presented after round 1	3.1	3.0	1.0	5.0	1.1
11. The impact data presented after rounds 1 and 2	3.6	4.0	1.0	5.0	1.1
12. The feedback report on estimated raw score cuts from rounds 1 and 2	3.6	4.0	1.0	5.0	1.1
13. Your interactions with your fellow panelists in your group before **round 1**	3.1	3.0	1.0	5.0	1.2
14. Your interactions with your fellow panelists in your group before **round 2**	3.2	3.0	1.0	5.0	1.2
15. Your interactions with your fellow panelists in your group before **round 3**	3.2	3.0	1.0	5.0	1.2
16. Your interactions with your fellow panelists in the large group discussion	3.1	3.0	1.0	5.0	1.1

Post-Policy Measurement Review Panel

The Post-Policy Measurement Review Panel met on the afternoon of Tuesday, June 23, after the completion of the item mapping committee meetings. The item mapping meeting and the Post-Policy Measurement Review Panel meeting were held in Albany. The Post-Policy Measurement Review Panel was convened with panelists from the Pre-Policy Measurement Review Panel. The purpose of the Post-Policy Measurement Review Panel was to integrate results from the Pre-Policy Measurement Review Panel meeting and the two committees from the item mapping meeting.

Panelists

This meeting was convened with the same thirty panelists from the Pre-Policy Measurement Review Panel.

Method and Procedure

The Post-Policy Measurement Review Panel meeting began with introductions of the facilitators and NYSED staff members. Panelists were then introduced to the purpose of the meeting. Panelists were instructed that they were to review and integrate results from the Pre-Policy Measurement Review and the item mapping. The product of this activity would be final recommendations for the percentage of students in each achievement level reflecting the influence of both meetings.

After these initial activities, panelists reviewed results from the Pre-Policy Measurement Review Panel. They were also given an explanation of the item mapping methodology. They then reviewed the results for committees A and B from the item mapping and synthesis meetings.

After the review of the methods and results of previous meetings, panelists were asked to try to independently integrate results from both meetings. They then discussed the integration of these results. Finally, the panelists made independent recommendations as to the percentage of students in each achievement level.

After these independent recommendations, the panelists were presented with the mean, median, minimum, and maximum percentage of students in each achievement level for the committee. They were asked to share with the rest of the committee how they integrated the results from the previous meetings.

Finally, panelists indicated that they favored some type of phase-in over years of the cut scores. Discussion followed. Panelists were asked to indicate their preference on the evaluation form. Only six out of the thirty panelists wrote

their preference on the evaluation form: four favored a phase-in plan, one did not favor the phase-in plan, and one did not indicate preference in either direction.

Results

Table 21 summarizes, for the Post-Policy Measurement Review Panel, the panelists' recommendations for the percentage of students who should be classified as Level 2 and above, and the percentage of students who should be classified as Level 3. Because there was a large outlier in the round one results, based on the requests from the panelists, the outlier was removed and the results were presented in Table 22. Table 23 presents the final round of recommendations from the post-policy meeting. The panelists requested that the median value be used for the overall recommendation due to outliers. Raw cut scores corresponding to the median recommendation from both rounds were presented to the panelists during the meeting.

Table 21. The Post-Policy Measurement Review Panel Results for the Recommended Percent of Students Who Should Be Classified as Level 2 and Above or Level 3, Round 1

	Level 2 and Above	Level 3
Mean	71	24
Median	73	25
Maximum	90	40
Minimum	20	10
Standard Deviation	13.7	6.5

Table 22. The Post-Policy Measurement Review Panel Results for the Recommended Percent of Students Who Should Be Classified as Level 2 and Above or Level 3, with Outlier Removed, Round 1

	Level 2 and Above	Level 3
Mean	73	24
Median	75	25
Maximum	90	40
Minimum	50	10
Standard Deviation	9.3	6.2

Table 23. The Post-Policy Measurement Review Panel Results for the Recommended Percent of Students Who Should Be Classified as Level 2 and Above or Level 3, Final Round

	Level 2 and Above	Level 3
Mean	72	24
Median	73	25
Maximum	90	40
Minimum	60	15
Standard Deviation	6.6	6.1

Figure 7 and Figure 8 show the percentage of students in each performance level using the mean and median recommendations from the Post-Policy Measurement Review Panel, with all data points. Figure 9 and Figure 10 present the percentage of students in each achievement level using the mean and median recommendations from the Post-Policy Measurement Review Panel, with an outlier being removed. Figure 11 and Figure 12 present the percentage of students in each achievement level using the mean and median recommendations for the final round.

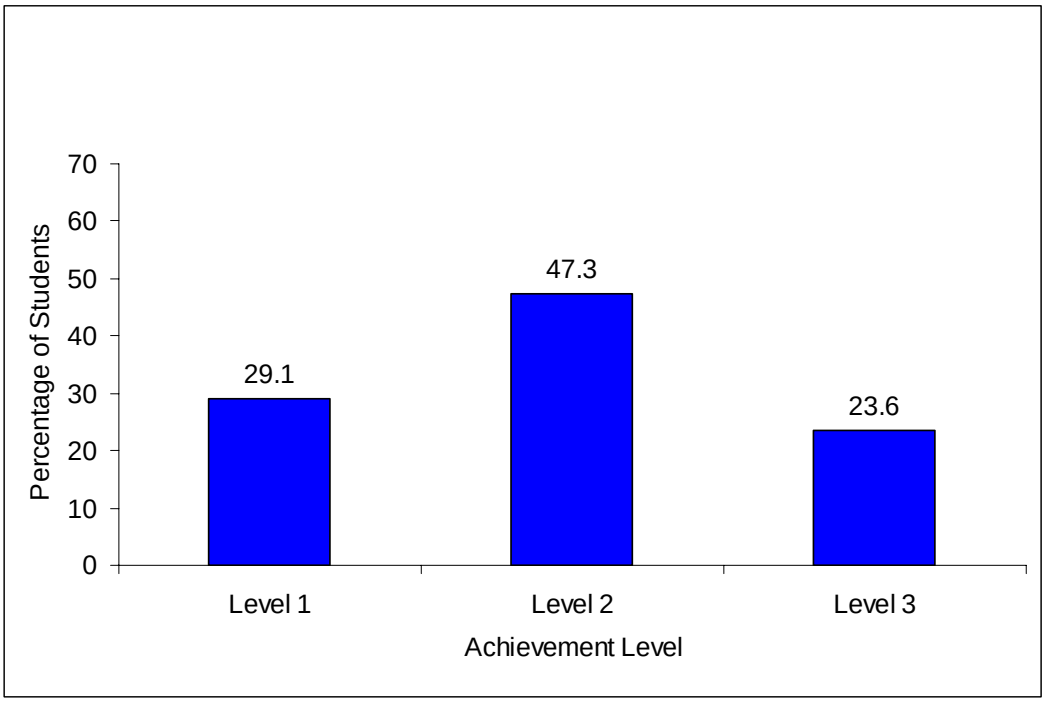

Figure 7. The Percentage of Students in Each Achievement Level Based on the Mean Recommendations from the Post-Policy Measurement Review Panel, Round 1

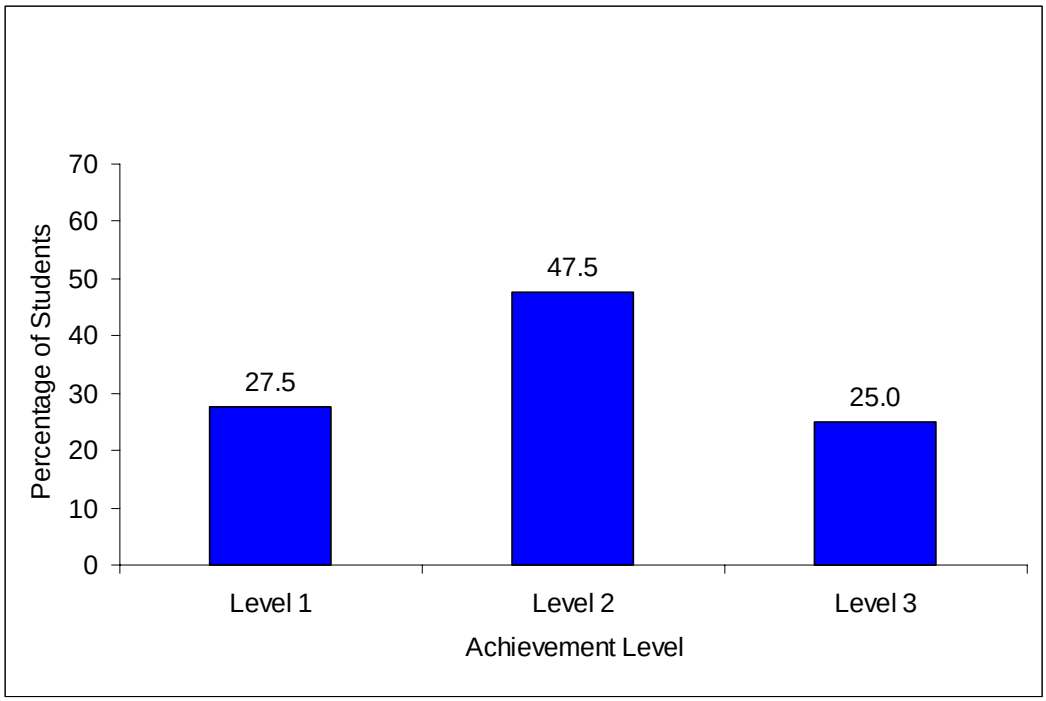

Figure 8. The Percentage of Students in Each Achievement Level Based on the Median Recommendations from the Post-Policy Measurement Review Panel, Round 1

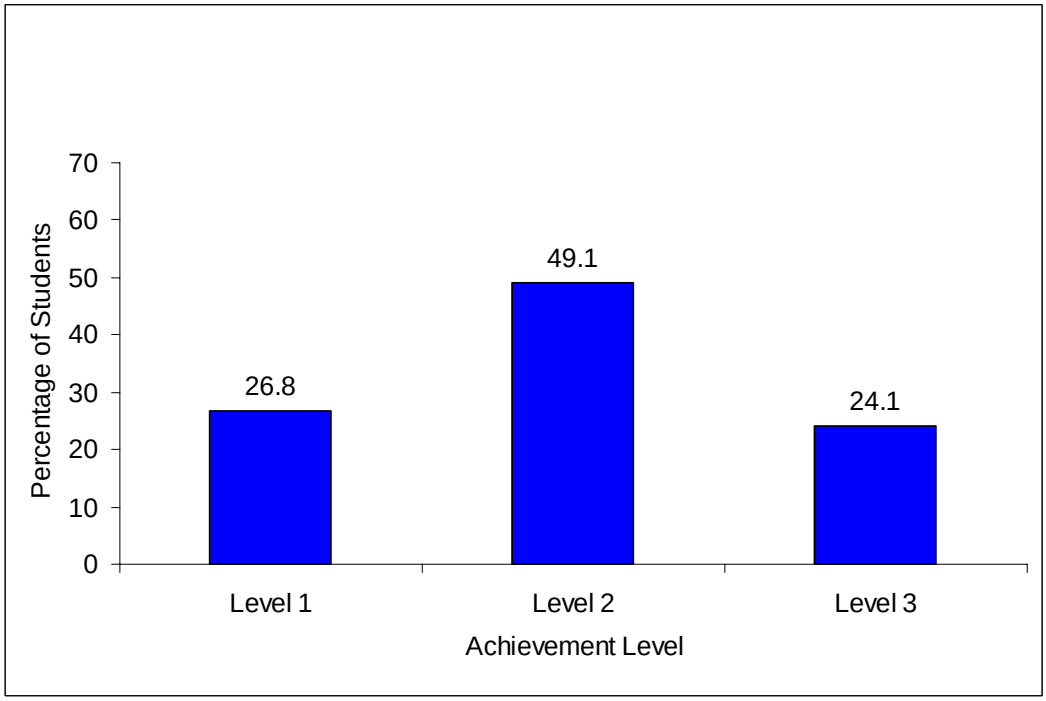

Figure 9. The Percentage of Students in Each Achievement Level Based on the Mean Recommendations from the Post-Policy Measurement Review Panel, with Outlier Removed, Round 1

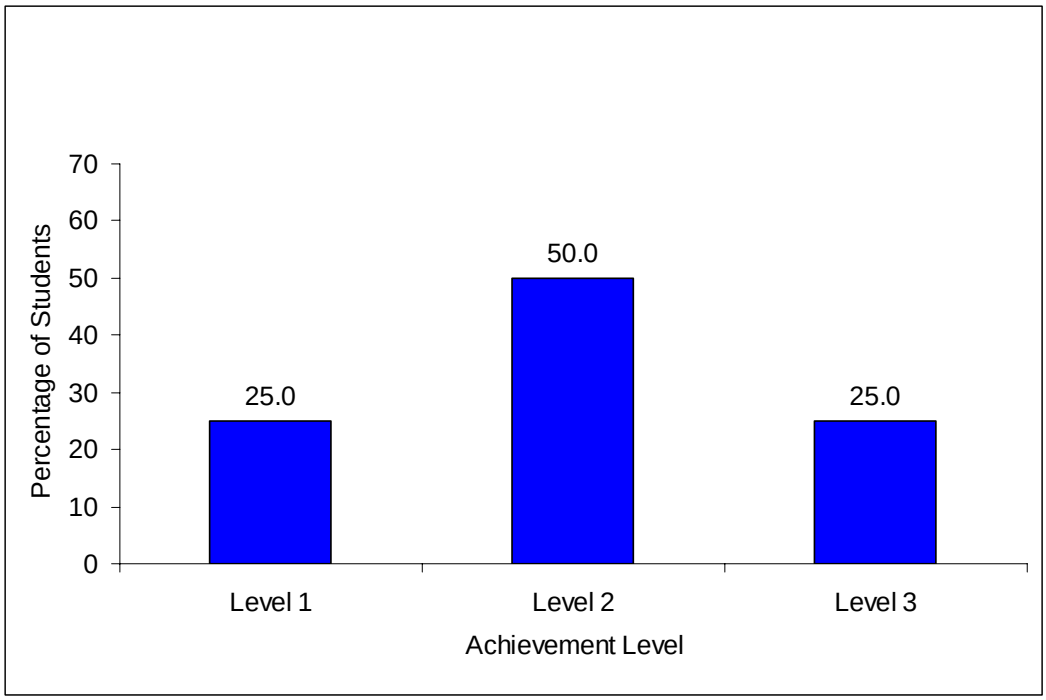

Figure 10. The Percentage of Students in Each Achievement Level Based on the Median Recommendations from the Post-Policy Measurement Review Panel, with Outlier Removed, Round 1

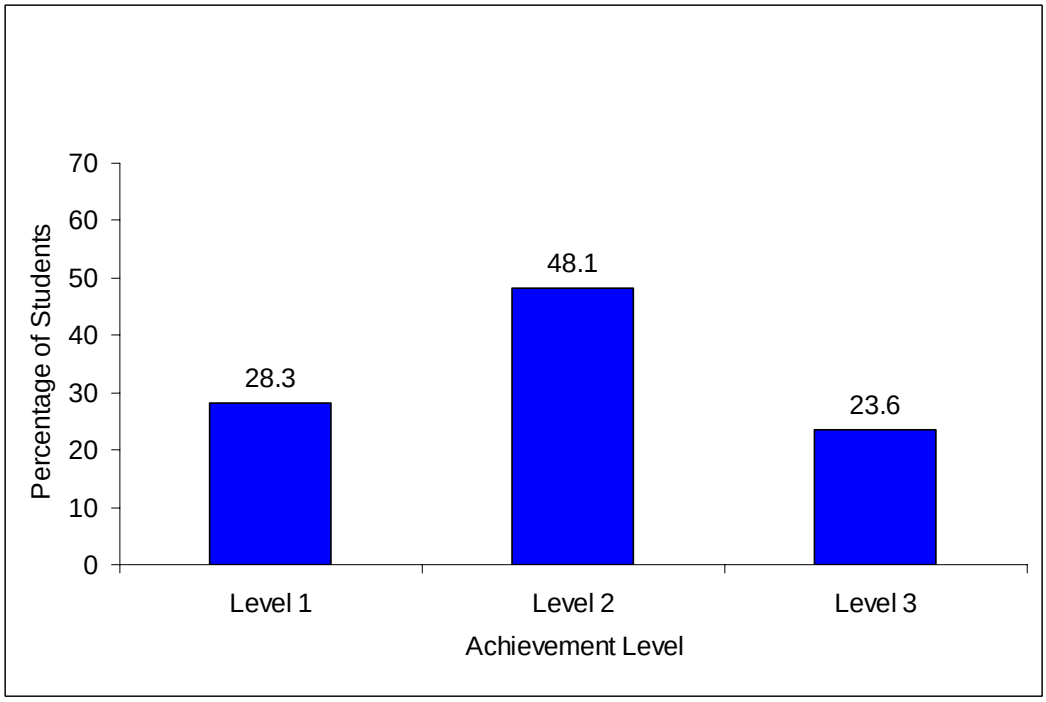

Figure 11. The Percentage of Students in Each Achievement Level Based on the Mean Recommendations from the Post-Policy Measurement Review Panel, Final Round

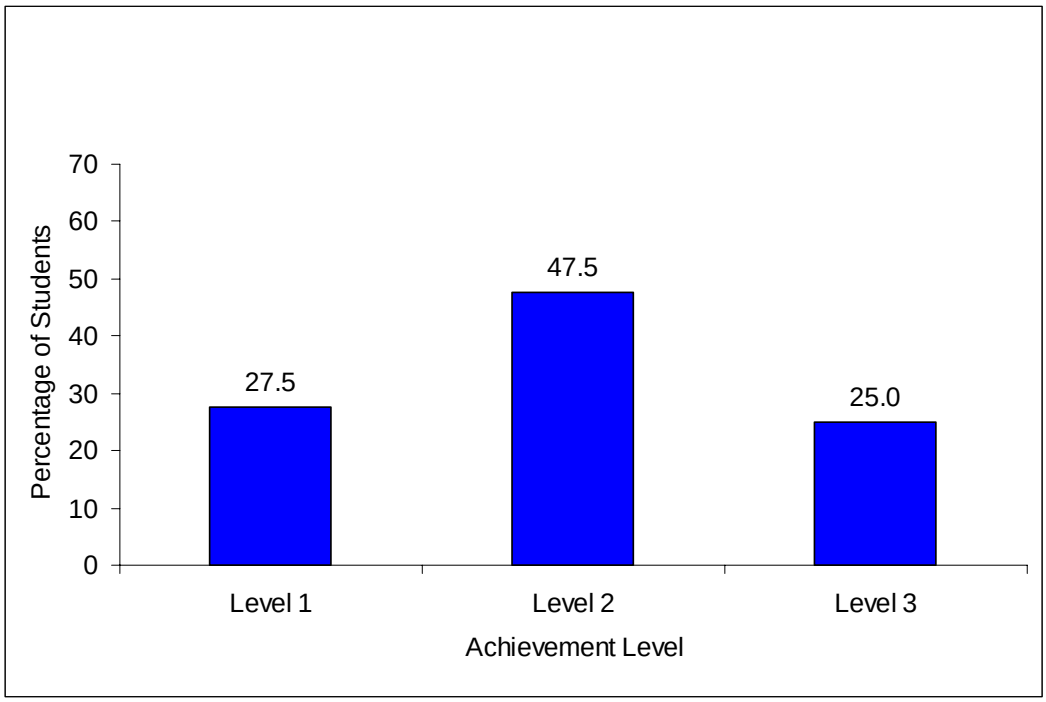

Figure 12. The Percentage of Students in Each Achievement Level Based on the Median Recommendations from the Post-Policy Measurement Review Panel, Final Round

Evaluations

An exit survey was completed by each panelist after the completion of the Post-Policy Measurement Review Panel meeting. Panelists answered each question using a scale of 1–5, 1 being "**totally disagree**" and 5 being "**totally agree**." The survey questions and the results are shown in Table 24.

Table 24. Questionnaire Results for the Post-Policy Measurement Review Panel

Question	Mean	Median	Minimum	Maximum	Standard Deviation
The method for making recommendations on the ideal percentage of students who should be classified in each achievement level was conceptually clear.	4.4	4.0	2.0	5.0	0.7
I had a good understanding of the results from the earlier meeting of Pre-Policy Measurement Review.	4.9	5.0	3.0	5.0	0.4
I had a good understanding of the results from the earlier Item Mapping Meeting.	4.8	5.0	2.0	5.0	0.6
After the first round of ratings, I felt comfortable with the method for making recommendations.	4.4	5.0	2.0	5.0	0.8
After the second round of ratings, I felt comfortable with the method for making recommendations.	4.6	5.0	3.0	5.0	0.6
I found the feedback on the recommendations of other panelists useful in making my second round recommendations.	4.3	4.0	3.0	5.0	0.5
I found the feedback on the overall group recommendation useful in making my second round recommendations.	4.2	4.0	2.0	5.0	0.8
I feel confident that the final cut score recommendations reflect the achievement levels associated with the New York State Regents Examination in Geometry.	4.0	4.0	1.0	5.0	1.0

Final Recommendation and Decision

As described in previous sections, NYSED conducted a formal standard setting process with Pearson that consisted of the following activities:

1) Pre-Policy Measurement Review
2) Item Mapping Standard Setting
3) Post-Policy Measurement Review

The three activities went according to plan and reflected the TAC's overview and recommended suggestions. The standard setting groups were diverse and representative of New York State. All groups adhered to instructions and processes put forward to them from the lead standard setting staff of Pearson. All activities were formally observed by the Office of State Assessment's Senior Managers and psychometric research staff.

After the standard setting activities, a conference call was set up between NYSED management and research staff members, TAC members, and Pearson psychometricians leading the standard setting meetings. The standard setting process and results were presented to the TAC, and the TAC formally endorsed NYSED's administration of the standard setting processes.

Below is a summary of the final standard setting recommendations from the content perspective and the policy perspective:

- Content perspective:
    - The raw cut score for a scale score of 65 was 50.
    - The raw cut score for a scale score of 85 was 71.
- Policy perspective:
    - The final percentage of students reaching Level 2 and above (with a scale score of 65 or higher) should be 72.5%, which, using impact data, translated to a raw score of 40[1].
    - The final percentage of students reaching Level 3 (with a scale score of 85 or higher) should be 25%, which, using impact data, translated to a raw score of 71

After careful consideration of such factors as the nature of the assessment, how rigorous the new curriculum is and how teachers in the field are adjusting to teach it, the role of the assessment in students' learning and advancement in high school, its desired impact, and so on, the senior management team made recommendations on the cut scores to the New York

[1] Due to observed outliers from the Post-policy Measurement Review Meeting, the committee required that their recommendations be calculated based on the median, instead of the mean of the ratings. Median is also the statistic used for the item mapping recommendations, as described in the report.

State Commissioner of Education. The Commissioner decided on the final cut scores for the Regents Examination in Geometry, and they are presented below:

- The final raw cut score for a scale score of 65 was 41, 1.5 standard error of measurement (SEM) below the item mapping final recommendation of 50.
- The final raw cut score for a scale score of 85 was 71, the same as the final recommended cut from the item mapping final recommendation.

The final impact of the cut scores on the students taking the Geometry assessment in June 2009 is as follows:

- 0–64 (Level 1), 28.5%
- 65–84 (Level 2), 48.2%
- 85–100 (Level 3), 24.4%

The SEM used for adjustment here was the SEM on the raw score scale of the overall test. For all Regents Examinations, scale scores are historically obtained by applying transformations to raw scores. Geometry is no exception. Therefore, it is reasonable that the SEM adjustment should also be based on raw score SEM. The following formula is used to determine the raw score SEM on the overall test

$SEM_{raw\,score} = SD_{raw\,score}\sqrt{1-\rho_{XX'}}$,

where $SD_{raw\,score}$ is the standard deviation of the raw scores on the overall test level and $\rho_{XX'}$ is the Cronbach alpha reliability estimate of the overall test. Using the census data collected from the June 2009 administration, the raw score mean was 53.422, and the raw score standard deviation was 19.618. The Cronbach alpha was 0.908. Substituting the values into the above formula, the SEM value was computed to be 5.96.

As the report described, the standard setting process was conducted carefully, and the best psychometric practices were followed. The policy decisions adhered to sound measurement principles to guarantee a thoughtful setting of cut scores, and NYSED is staying consistent with the approaches that have been integrated to the State's standard setting processes and that have been used with the Grades 3–8 Testing Program and the Regents Examination in Integrated Algebra.

Appendix A

Agenda for the Pre-Policy Measurement Review Panel Meeting

University of the State of New York
State Education Department

PEARSON

Agenda for Pre-Policy Measurement Review Meeting
Regents Examination in Geometry
April 9, 2009

Registration	7:30-8:00
Welcome and Introductions	8:00-8:30
Purpose of Meeting	
Introductions	
Review of Agenda	
Non-disclosure Agreement Form	
Reimbursement Forms	
Opening Remarks	8:30-9:00
David Abrams, Assistant Commissioner	
Overview of Measurement Review	9:00-10:00
Purpose	
Methodology	
BREAK	10:00-10:15
Presentation of Related Assessments	10:15-11:00
New York State Grade 8 Mathematics Test	
Regents Examination in Mathematics A	
Regents Examination in Mathematics B	
Regents Examination in Integrated Algebra	
National Assessment of Education Progress (NAEP)	
Break into Pre-assigned Groups	11:00
Discuss Assessment Data and Guiding Questions	11:00-11:45
Round 1 Recommendations	11:45-12:00
Readiness Form	
Review Method	
Collect Recommendations	
LUNCH	12:00-1:00
Large Group Feedback	1:00-2:00
Break into Pre-assigned Groups	2:00
Discuss Feedback in Small Groups	2:00-2:30
Round 2 Recommendations	2:30-2:45
Readiness Form	
Review Method	
Collect Recommendations	
BREAK	2:45-3:00
Feedback on Final Recommendations	3:00-3:15
Complete Survey	3:15-3:30
End of Day Activities (Check in Materials)	3:30-3:45

Appendix B

Recommendation Form for the Pre-Policy Measurement Review Panel Meeting

Rating Form

REGENTS EXAMINATION IN GEOMETRY
PRE-POLICY MEASUREMENT REVIEW
RATING FORM

Directions: Your task is to recommend the ideal percent of students who should be classified in each performance level on the Regents Examination in Geometry. First, recommend the ideal percent of students who should be classified as *Level 2 and above*. Next, recommend the ideal percent of students who should be classified as *Level 3*. Make your recommendations using values that are multiples of 5, for example, 50, 55 or 60.

Panelist ID: _____

Table: _____

Round 1

Low End of Range	
% Level 2 & Above	% Level 3
High End of Range	
% Level 2 & Above	% Level 3

Round 2

Low End of Range	
% Level 2 & Above	% Level 3
High End of Range	
% Level 2 & Above	% Level 3

Appendix C

Demographic Questionnaire for the Item Mapping Committee Meetings

Pre-Policy Measurement Review
Regents Examination in Geometry
Panelist Information Sheet

Panelist ID: _____

Please provide the following demographic information that will be used to describe the general characteristics of the panelists who are recommending standards as members of the Pre-Policy Measurement Review Committee for the Regents Examination in Geometry.

Your Current Position:

Gender (circle one): Male Female

Ethnicity:

Years of Educational Experience:

Compared to other school districts in New York State, how would you describe the size of your district (circle one)?

Large Medium Small

Compared to other school districts in New York State, how would you describe the location of your district (circle one)?

Urban Suburban Rural

Compared to other school districts in New York State, how would you describe the geographic location of your district (circle one)?

North South East West Central

Appendix D

Agenda for the Item Mapping Committee Meeting

University of the State of New York
State Education Department

David Abrams
Assistant Commissioner
Office for Standards, Assessments and Reporting

PEARSON

2510 North Dodge Street
Iowa City, IA 52245
(800) 627-7990

Recommendations for Setting Achievement Levels for the Regents Examination in Geometry
Agenda

DAY 1-June 22, 2009

Registration	7:30-8:00
Welcome and Opening Remarks – Paul Nichols Introductions Why are you here? Agenda Non-disclosure forms Reimbursement forms	8:00-8:45
Welcome and Overview – David Abrams Examination Development Overview of Standard Setting Process	8:45-9:15
Standard Setting Process – Paul Nichols Purpose Item Mapping Methodology	9:15-9:45
Overview of the Regents Examination in Geometry History Purposes Test specifications	9:45-10:00
BREAK	10:00-10:15
Introduce Achievement Level Descriptors	10:15-10:30
Construct Achievement Level Descriptors Small Group Discussion	10:30-12:00
LUNCH Train Table Leaders	12:00-1:00

Construct Achievement Level Descriptors Large Group Discussion	1:00-1:30
Break to two committee rooms Introduction	1:30-1:45
Complete Selected Geometry Items on the Examination	1:45-2:45
BREAK Assign Panelist ID's	2:45-3:00
Overview of Standard Setting Item Mapping Ordered Item Booklet Item Map Ratings Forms	3:00-3:30
Practice Round	3:30-3:45
Round 1 Standard Setting Readiness Form Review Method Collect Completed Rating Forms	3:45-4:45
End of Day Activities Review Day 2 Schedule Check in Materials	4:45-5:00

END OF DAY 1

DAY 2 – June 23, 2009

Registration	8:00-8:15
Review Schedule, Answer Questions (both committees)	8:15-8:30
Reconvene in Committee Rooms	8:30
Feedback Small Group Discussion of Table Agreement Data	8:30-9:15
Round 2 Ratings Readiness Form Review Method Collect Completed Rating Forms	9:15-10:15
BREAK	10:15-10:45
Feedback Small Group Discussion of Table Agreement Data Committee Discussion of Group Agreement Data Committee Discussion of Impact Data	10:45-11:30
Round 3 Ratings Readiness Form Review Method Collect Completed Rating Forms	11:30-12:00
LUNCH	12:00-1:00
Feedback	1:00-1:15
Revisit Achievement Level Descriptors Revise Achievement Level Descriptors	1:15-2:00
Closing Remarks – David Abrams	2:00-2:15
Complete Survey	2:15-2:30
End of Day Activities Check in Materials	2:30-3:00

University of the State of New York
State Education Department

David Abrams
Assistant Commissioner
Office for Standards, Assessments and Reporting

PEARSON

2510 North Dodge Street
Iowa City, IA 52245
(800) 627-7990

Agenda for Standard Setting Synthesis Group Discussion
New York State Regents Examination in Geometry
June 23, 2009

Overview of the Synthesis Review	3:00-3:10
Purpose	
Methodology	
Presentation from Each Committee	3:10-3:30
Presentation from Committee A	
Presentation from Committee B	
Group Discussion	3:30-4:30
Focus on Items Around the Cuts	
Focus on Skills and Knowledge	
Rating	4:30-5:00

Appendix E
Achievement Level Descriptors

New York State Regents Examination in Geometry
Achievement Levels

- **Not Passing**
  - A ***not passing*** student is unable to demonstrate, on demand, proficiency in understanding the content and concepts required for commencement-level achievement in any or most of the learning standards and key ideas assessed.
  - A ***not passing*** student is unable to demonstrate on demand, proficiency in the skills required for commencement-level achievement in any or most of the learning standards and key ideas assessed.
  - A ***not passing*** student is unable to demonstrate, on demand, evidence of an ability to apply the content, concepts, and skills required to meet any or most of the demands of productive adult citizenship, the workplace, and postsecondary education.

- **Passing**
  - A ***passing*** student is able to demonstrate, on demand, knowledge of the content and concepts required for commencement-level achievement in each of the learning standards and key ideas assessed.
  - A ***passing*** student is able to demonstrate, on demand, the skills required for commencement-level achievement in each of the learning standards and key ideas assessed.
  - A ***passing*** student is able to apply, on demand, the content, concepts, and skills required to meet the demands of productive adult citizenship, the workplace, and postsecondary education.

- **Passing with Distinction**
  - A ***passing with distinction*** student is able to demonstrate, on demand, evidence of superior understanding of the content and concepts required for commencement-level achievement in each of the learning standards and key ideas assessed.
  - A ***passing with distinction*** student is able to demonstrate, on demand, evidence of superior skills required for commencement-level achievement in each of the learning standards and key ideas assessed.
  - A ***passing with distinction*** student is able to demonstrate, on demand, evidence of superior ability to apply the content, concepts, and skills required to meet the demands of productive adult citizenship, the workplace, and postsecondary education.

New York State Regents Examination in Geometry
Learning Standard

Mathematics, Science, and Technology - Standard 3

Students will:

- understand the concepts of and become proficient with the skills of mathematics;
- communicate and reason mathematically;
- become problem solvers by using appropriate tools and strategies;
- through the integrated study of number sense and operations, algebra, geometry, measurement, and statistics and probability.

New York State Regents Examination in Geometry
Achievement Level Descriptors
June 22–23, 2009

- Description of Students at Level 2 (65–84)
  - Engaged and present most of the time
  - Take responsibility for learning
  - Have a foundation of basic algebra skills and an ability to solve equations (e.g., manipulate the equation of a line but may not be able to apply it)
  - Apply algebra skills with geometric concepts, but not necessarily correctly
  - College-ready with minimal remedial courses
  - Knowledge of basic definitions, most vocabulary (e.g., centroid)
  - Vocabulary and reading comprehension are good, not stellar; can read and understand, distinguish what the problem is about and know what is to be solved
  - Able to make a connection to basic, real-world situations and draw from prior knowledge (courses, real life, etc.)
  - Able to reason and draw logical conclusions
  - Can set up an equation, label and extract information from a diagram but not necessarily solve
  - Show minor conceptual errors, understand little mistakes
  - Should know an answer makes sense
  - On constructed response problems, get 2 out of 4 or 4 out of 6 points
  - Show work
  - Know how to use formulas and tools supplied, know some symbols
  - Able to formulate some kind of argument
  - Recognize there is more than one way to solve a problem, but are only able to use one approach
  - Can begin a proof, show steps for simple proofs, apply and synthesize basic proofs, 3 or 4 out of 6
  - Recognize, comprehend, and apply geometric concepts and basic properties; may have a problem retaining concepts long term
  - Calculator-proficient, not dependent
  - Basic test-taking skills, plugging in/working backward/showing work/writing official answer/drawing pictures
  - When proving triangles, know the concepts of congruence and similarity, and are able to differentiate between types of triangles
  - Properties of different shapes, pictures help
  - Slopes, parallel and perpendicular lines with pictures
  - Construct a bisector
  - Can solve basic loci problems

- Pick out points of intersection
- Solve graphically, if y = type
- Able to work with circles in a limited capacity, such as arcs and angles
- Transformations: understand what they are and the differences between them; able to label (one step, not two)

- Description of Students at Level 3 (85−100)
  - Take notes, participate in class, study, seek help, ask questions, strive for success
  - Consistent, abstract thinkers, risk takers
  - Mastery of basic algebra skills
  - Incorporate algebra with geometry
  - Mastery of geometric relationships application, apply geometric notation and units of measurement, do radicals correctly
  - Use tools appropriately
  - Able to eliminate obviously wrong answers
  - Always check answers
  - Know most or all vocabulary, formulas, and appropriate symbols
  - Recognize the need for a particular theorem and apply it to complete a multi-step problem
  - Multiple problem-solving approaches, can solve problems in multiple ways
  - Transfer of knowledge to unique problems
  - Consistent application of concepts
  - Articulate justifications, able to explain work
  - Deep enough understanding to help others solve the problem
  - Few conceptual errors, maybe a computational error
  - May enter and solve the problem differently from other students
  - Work with or without a calculator
  - Able to completely work through most proofs
  - Valid in application and solving of proofs
  - Apply geometric shapes, such as perpendicular and parallel lines; apply knowledge of slopes and perpendicular lines
  - Know what constructions mean
  - Able to sketch compound loci
  - Compose transformations
  - Use Pythagorean theorem

- Description of Students at Level 1 (0−64)
  - Bored, less motivated
  - Easily frustrated
  - Variety of problems with attendance, motivation, and confidence
  - May have many external issues affecting the ability to learn
  - Need constant reinforcement

- Did not have prior success, therefore, come in with preconceived knowledge
- Weak English skills
- Limited recognition of vocabulary and reading comprehension problems
- Poor or delayed processing skills, comprehension is poor, unable to translate from English, unable to retrieve information, poor auditory comprehension
- Recognize previously seen vocabulary but don't know the meaning
- Don't read whole questions or won't answer the questions that were asked; don't show work
- Sometimes have better calculator skills because of heavy reliance on calculators, although not necessarily
- Don't understand the validity of answers; rely heavily on a calculator for basic math skills
- Gaps in knowledge and prior skills, may not have passed prerequisite courses, limited algebra skills
- Have difficulty making connections to prior knowledge
- May have taken Algebra for two years
- Lack of retention, cannot recall a year's worth of information or information from previous courses
- Make computational errors
- Show little or no work on problems, or leave blank
- With problem solving, don't have a clear plan on how to solve a problem, can recopy things from a problem
- Recognition of formulas, but can't use them
- Put givens into a proof, but can't solve
- Repetition is necessary
- Memorize instead of understand concepts
- Know that triangles are congruent, but cannot prove why
- If told something is midpoint, can write a statement that it is the midpoint, but nothing else
- Need visuals
- Can get "what" questions correct
- Can do individual topics, but cannot tackle multiple steps or synthesize information
- Can work two-dimensional problems, but not three-dimensional problems
- Have difficulty visualizing concepts or reading diagrams
- Some symbol use
- Can write a few angles on a proof, but not much else
- Know certain shapes and properties, but cannot explain them
- Can calculate basic area, but don't know how to apply it
- Can plot points, but stop there

- Most Distinguishing Characteristics of Students at the Threshold between Level 2 (65−84) and Level 1 (0−64)
    - Can start a problem
    - Can apply a formula that is already given
    - Solve a problem using algebraic skills, but will not understand concepts behind their use
    - Demonstrate and apply problem-solving skills limited to 1 or 2 steps
    - Extract some relevant information from a problem
    - Will show more work, will set up a problem correctly but get lost in the sequencing (e.g., can set up 2 columns, set up the given, know that what is to be proven is last but may not know all the steps in between)
    - Have better vocabulary skills
    - Know basic geometry
    - Test-taking skills are better, use test-taking strategies such as working backward
    - Have slightly better Algebra skills
    - Can get recall questions correct, but not the application
    - Have trouble with a formula sheet
    - Have problem-solving skills

- Most Distinguishing Characteristics of Students at the Threshold between Level 3 (85−100) and Level 2 (65−84)
    - More successful on open-ended questions
    - Work presentation is of high quality, show a lot of work
    - Independent learners; will study formulas to try to understand the concepts
    - Ability to make connections between different concepts
    - Ability to check work and fix errors
    - Know more than one way to attempt/approach problems
    - Have better organized thought processes
    - Planned problem solving
    - Check reasonableness
    - Have time management skills
    - Correctly understand a concept but make computational errors
    - Know the flow that the proof should follow, but maybe not all the reasons for every step
    - Have better retention and memorization skills
    - Have stronger logic skills
    - Make fewer computational and minimal conceptual errors
    - Are a little more careful or meticulous of work
    - Attempt all and finish most questions
    - Work outside of class
    - Take roundabout way to get answers, but not necessarily how it was taught
    - More consistent with algebraic applications

- Able to sketch diagrams for a visual
- Have a higher level of reading comprehension
- Consistently draw diagrams and correctly label them; complete an accurate diagram from a description
- Are able to think abstractly
- Tends to know the way to solve a problem
- Can solve problems and explain more steps to getting the answer

Appendix F

Recommendation Form for the Item Mapping Committee Meetings

New York State Regents Examination in Geometry
Item Position Recording Sheet

Panelist ID _____

		Round 1	Initials		Round 2	Initials		Round 3	Initials
65	**Recommended Cut**								
85	**Recommended Cut**								

Appendix G

Agenda for the Post-Policy Measurement Review Committee Meeting

University of the State of New York
State Education Department

David Abrams
Assistant Commissioner
Office for Standards, Assessments and Reporting

PEARSON

2510 N Dodge Street
Iowa City, IA 52245
(800) 627-7990

New York State Regents Examination in Geometry
Post-Policy Measurement Review
Agenda

Registration	4:30-4:45
Dinner	4:45-5:30
Welcome and Overview – David Abrams Introductions Overview of Standard Setting Process	5:30-5:45
Overview of Post-Policy Measurement Review Logistics Purpose Methodology Present Results from Standard Setting Discussions	5:45-6:30
Round 1 Recommendations Readiness Form Review Method Collect Recommendation	6:30-6:45
Break	6:45-7:00
Discussions Round 1 Results Group Discussion	7:00-7:20
Round 2 Recommendations Readiness Form Review Method Collect Recommendation	7:20-7:30

Appendix H

Recommendation Form for the Post-Policy Measurement Review Committee Meeting

NEW YORK STATE REGENTS EXAMINATION IN GEOMETRY
POST-POLICY MEASUREMENT REVIEW
RATING FORM

Directions: Your task is to recommend the ideal percentage of students who should be classified in each achievement level on the Regents Examination in Geometry. First, recommend the ideal percentage of students who should be classified as *Level 2 and above*. Next, recommend the ideal percentage of students who should be classified as *Level 3*. Make your recommendations using values that are multiples of 5, for example, 50, 55 or 60.

Please recommend a range of the ideal percentage of students for both Level 2 and above, and Level 3.

Panelist ID: _____

Round 1

Level 2 and Above	
Level 3	

Round 2

Level 2 and Above	
Level 3	

Appendix I

Ordered Item Book

Table I1. Ordered Item Book

Page Number	Item ID	θ with RP 0.67	Page Number	Item ID	θ with RP 0.67
1	OPITEM_34_1	-0.791	46	OPITEM_12	0.973
2	OPITEM_2	-0.697	47	OPITEM_38_3	1.023
3	OPITEM_34_2	-0.411	48	OPITEM_11	1.033
4	OPITEM_13	-0.327	49	ANCHOR2_3	1.053
5	ANCHOR1_2	-0.327	50	OPITEM_32_1	1.067
6	ANCHOR1_9_1	-0.282	51	OPITEM_35_4	1.071
7	OPITEM_4	-0.237	52	ANCHOR1_7	1.143
8	OPITEM_25	-0.227	53	OPITEM_30_1	1.145
9	OPITEM_5	-0.157	54	OPITEM_15	1.223
10	OPITEM_1	-0.047	55	ANCHOR2_10_2	1.224
11	ANCHOR1_9_2	-0.018	56	OPITEM_37_3	1.23
12	OPITEM_8	0.143	57	OPITEM_38_4	1.261
13	OPITEM_10	0.163	58	OPITEM_26	1.323
14	OPITEM_38_1	0.204	59	OPITEM_29_2	1.362
15	OPITEM_3	0.223	60	ANCHOR2_6	1.373
16	ANCHOR2_4	0.223	61	OPITEM_33_1	1.43
17	ANCHOR2_1	0.253	62	OPITEM_36_3	1.432
18	ANCHOR2_5	0.293	63	OPITEM_18	1.443
19	OPITEM_37_1	0.318	64	OPITEM_37_4	1.563
20	OPITEM_6	0.363	65	OPITEM_23	1.603
21	OPITEM_7	0.423	66	ANCHOR1_1	1.653
22	ANCHOR2_2	0.443	67	OPITEM_31_2	1.822
23	OPITEM_35_1	0.481	68	OPITEM_30_2	1.841
24	ANCHOR2_7	0.483	69	OPITEM_38_5	1.849
25	OPITEM_29_1	0.518	70	OPITEM_24	1.873
26	OPITEM_22	0.523	71	OPITEM_32_2	1.943
27	OPITEM_17	0.573	72	ANCHOR1_5	1.993
28	ANCHOR1_3	0.653	73	OPITEM_33_2	2.013
29	ANCHOR1_6	0.653	74	ANCHOR1_10_1	2.257
30	OPITEM_38_2	0.666	75	ANCHOR2_8_1	2.267
31	ANCHOR2_9_1	0.67	76	ANCHOR2_8_2	2.383
32	OPITEM_37_2	0.68	77	ANCHOR2_8_3	2.383
33	OPITEM_21	0.683	78	OPITEM_27	2.463
34	ANCHOR2_10_1	0.685	79	ANCHOR1_10_2	2.474
35	OPITEM_35_2	0.692	80	OPITEM_38_6	2.531
36	OPITEM_19	0.703	81	OPITEM_28	2.613
37	OPITEM_36_1	0.727	82	ANCHOR1_8_1	2.705
38	OPITEM_16	0.793	83	ANCHOR1_10_3	2.748
39	OPITEM_14	0.793	84	ANCHOR2_9_2	2.823
40	OPITEM_9	0.813	85	ANCHOR1_10_4	2.939
41	OPITEM_35_3	0.851	86	ANCHOR1_8_2	3.185
42	ANCHOR1_4	0.933	87	ANCHOR1_10_5	3.198
43	OPITEM_36_2	0.961	88	OPITEM_36_4	3.389
44	OPITEM_31_1	0.962	89	ANCHOR1_10_6	3.652
45	OPITEM_20	0.963	90	ANCHOR2_8_4	5.933

Table I2. Raw to θ Conversion Table

Raw Score	θ	CSEM	Raw Score	θ	CSEM
0	-5.021	1.831	44	0.410	0.208
1	-4.308	1.010	45	0.453	0.207
2	-3.595	0.722	46	0.495	0.206
3	-3.169	0.595	47	0.538	0.206
4	-2.860	0.520	48	0.580	0.205
5	-2.617	0.469	49	0.622	0.205
6	-2.414	0.433	50	0.664	0.206
7	-2.240	0.404	51	0.707	0.206
8	-2.085	0.382	52	0.749	0.207
9	-1.947	0.363	53	0.792	0.208
10	-1.821	0.348	54	0.836	0.209
11	-1.705	0.335	55	0.880	0.210
12	-1.597	0.323	56	0.924	0.212
13	-1.495	0.314	57	0.970	0.214
14	-1.399	0.305	58	1.016	0.216
15	-1.309	0.298	59	1.063	0.219
16	-1.222	0.291	60	1.111	0.221
17	-1.139	0.285	61	1.161	0.225
18	-1.059	0.280	62	1.212	0.228
19	-0.983	0.275	63	1.265	0.232
20	-0.908	0.270	64	1.320	0.236
21	-0.836	0.266	65	1.376	0.240
22	-0.767	0.262	66	1.435	0.245
23	-0.699	0.259	67	1.497	0.251
24	-0.633	0.255	68	1.561	0.257
25	-0.569	0.252	69	1.629	0.263
26	-0.506	0.249	70	1.700	0.270
27	-0.445	0.246	71	1.775	0.278
28	-0.385	0.243	72	1.855	0.287
29	-0.327	0.240	73	1.940	0.297
30	-0.270	0.237	74	2.032	0.308
31	-0.215	0.234	75	2.130	0.321
32	-0.161	0.231	76	2.238	0.335
33	-0.108	0.229	77	2.356	0.352
34	-0.056	0.226	78	2.487	0.372
35	-0.005	0.224	79	2.635	0.397
36	0.044	0.221	80	2.804	0.426
37	0.093	0.219	81	3.002	0.465
38	0.140	0.217	82	3.242	0.517
39	0.187	0.215	83	3.547	0.593
40	0.233	0.213	84	3.973	0.721
41	0.278	0.212	85	4.687	1.011
42	0.323	0.210	86	5.401	1.832
43	0.366	0.209			