New York State Regents Examination in

Geometry (Common Core)

Standard Setting Technical Report

Prepared for the New York State Department of Education

by

Data Recognition Corporation

July 20, 2015

Page intentionally left blank

Table of Contents

List of Tables

List of Figures

Executive Summary

The standard setting process for the New York State Regents Examination in Geometry (Common Core) consisted of two events: the Performance Level Description Development meeting and the standard setting meeting. The primary goal for these meetings was to establish cut scores that operationally define the five performance levels: Level 1, Level 2, Level 3, Level 4, and Level 5. The performance level designations will be used by local, state, and federal accountability programs and are central to communicating with parents, teachers, and the public. This document provides a detailed description of the activities held at each meeting.

On February 18, 2015, the Performance Level Description Development meeting was conducted in Albany, New York. The focus of this meeting was the development of performance level descriptions (PLDs), which describe the specific knowledge and skills of students at each level of performance. Each PLD was designed to describe the range of students at that performance level and was used in the subsequent standard setting meeting.

On June 16 and 17, 2015, a standard setting meeting was held. The purpose of this meeting was to identify four cut scores that distinguish the five levels of performance on the New York State Regents Examination in Geometry (Common Core). Using the PLDs articulated in February, panelists followed the Bookmark standard setting procedure, which resulted in cut score recommendations that were brought to the New York State Education Department (NYSED).

In this technical report, panelists, materials, methodologies, and results for each meeting are presented for the New York State Regents Examination in Geometry (Common Core). A preliminary summary of standard setting activities was presented to the NYSED the day following the standard setting meeting. This report provides final results and additional details documenting the standard setting process and the results.

Performance Level Description Development Meeting

On February 18, 2015, the New York State Education Department (NYSED) conducted a Performance Level Description Development meeting in Albany, New York. The meeting was convened to articulate the knowledge and skills expected of students at each level of performance, consistent with the policy vision set forth by the NYSED.

At this meeting, panelists were asked to review policy guidelines and content standards to generate knowledge and skill statements that describe a student at a specific level of performance. An initial training session regarding the overall process of standard setting and the specific role that PLDs play within standard setting was provided by the lead psychometrician. Using NYSED-approved training materials, educators were trained to deconstruct the content standards in terms of cognitive complexity and then to align these different aspects of the content standards with specific levels of performance. Educators were trained to adhere carefully to the cognitive alignment (e.g., depth of knowledge, cognitive complexity, and range of skills) with the State's content standards while keeping the policy decisions in mind.

Panelists

A total of 15 Geometry subject-matter expert educators attended the meeting. The participants were recruited by the NYSED.

Method and Procedure

The PLD meeting began with introductions of NYSED staff and the Data Recognition Corporation (DRC) facilitators. The lead DRC psychometrician provided an opening training session that included an overview of standard setting and the process by which cut scores are determined. The policy decisions associated with the standard setting, including the number of performance levels (five) and the associated labels for these levels, were reviewed. An explanation of how the PLDs document the expected knowledge and skills associated with each performance level was then provided. The role that the PLDs play in establishing the cut scores between each level of performance was described.

DRC content experts then described the development of the PLDs, which would use the four-step process described below.

Step 1. Review and Internalize Policy PLDs

Panelists reviewed the statements that describe the policy vision that the NYSED has for the Regents Examination performance levels. Panelists were instructed to use

this vision as the context for preparing the PLDs. Throughout the day, DRC facilitators reminded participants to recall the policy expressed in these statements.

Step 2. Evaluate Content Standards in Terms of Cognitive Complexity
DRC facilitators walked panelists through several examples of how to deconstruct the content standards in terms of the different levels of cognitive complexity until all panelists were oriented to this task. The content standards consist of statements that describe knowledge, skills, and performance, which range in terms of cognitive complexity; these statements are similar to the kinds of statements that are frequently included in PLDs. More importantly, the content standards include statements of basic skills that students would display, as well as other skills that require more advanced cognitive processing by students. The articulation of different levels of cognitive complexity reflected in content standards provides the basis for the development of the PLDs.

Step 3. Align Levels of Cognitive Complexity with Performance Levels
After a thorough evaluation of each content standard, participants identified specific statements that described different levels of cognitive complexity for various knowledge and skills specified within the content standards. Participants then classified each of these statements in terms of the different performance levels. That is, each statement that expressed some level of cognitive complexity within a content standard was categorized into the different performance levels (i.e., Level 1, Level 2, Level 3, Level 4, or Level 5). Some statements were clearly aligned with a given performance level. Other aspects of the content standards did not fall cleanly into a specific PLD; these reflected a transition point from one performance level to another.

DRC facilitators showed participants how to use a coding scheme to reflect which statements clearly aligned with specific levels of performance and which statements reflected transition points.

Step 4: Prepare Draft PLDs
The classification of the cognitive complexity of the content standards and associated skills in terms of the performance levels provided a straightforward framework that enabled participants to create initial drafts of the Range and Threshold PLDs. Skills from the content standards that were clearly associated with a specific performance level provided insight into what constituted the performance level for all students in that level (i.e., Range PLD). Similarly, the skills that spanned adjacent performance levels and were difficult to categorize provided insight into what constituted the transition between levels of performance (i.e., Threshold PLD, described below).

Upon completion of the subject-specific training, each group of panelists was divided into small teams. Each team was assigned several content standards. Participants

then deconstructed each content standard and identified the different statements about students being made in the standard in terms of cognitive complexity. Particular aspects of the content standards that were easily classified into a given performance level essentially formed the basis of the PLDs. Working in these teams, participants produced drafts of the PLDs. At the conclusion of this activity, the draft PLDs were shared across teams for cross-team discussion and revision. When teams encountered knowledge and skills that were difficult to classify into a particular level of performance, panelists were asked to document such challenges as potentially indicative of transitional knowledge and skills that demarcated the threshold between performance levels.

The drafts produced represented the participants' conceptualization of the range of students in each performance level. After the meeting, DRC, working with the NYSED, reviewed and revised the PLDs for clarity and consistency. The end result of this meeting was a set of PLDs that clearly defined the level of knowledge and skill necessary for each performance level.

Results

Copies of the final Geometry PLDs developed at this meeting and revised by DRC and the NYSED are provided in Appendix A. These PLDs were used in the subsequent standard setting meeting.

Bookmark Standard Setting Meeting

A committee of New York State educators was convened on June 16–17, 2015, in Albany, New York, to recommend performance standards for the New York State Regents Examination in Geometry (Common Core). The Geometry committee consisted of 28 educators.

DRC followed a Bookmark procedure similar to the method originally defined by Lewis, Mitzel, and Green (1996). The Bookmark procedure is arguably the standard setting method that is most philosophically consistent with criterion-referenced, standards-based assessments like the Regents Examinations. This method is discussed in detail within the Methods section of this document.

Panelists

All panelists (committee members) voluntarily provided demographic information. Five table leaders for were identified from the pool of panelists by the NYSED and DRC. Table 1 provides a summary of gender representation of panelists. Table 2 presents a summary of ethnic representation of panelists. Table 3 provides a geographic summary of panelists. Table 4 provides a summary of the educational background of the committee.

Table 1. Number of Male and Female Panelists in Committees

	Geometry
Female	17
Male	11

Table 2. Ethnic Composition of the Panelists in Committees

	Geometry
White	21
Hispanic	2
African American	1
Asian	2
Missing Information	2

Table 3. Geographic Locations of Panelists for Standard Setting

	Geometry
Big 4 Cities	2
Capital Region	3
Central NY	4
Hudson Valley	4
Long Island	2
NYC	6
Western NY	6
Southern Tier	1

Table 4. Education Roles of Panelists for Standard Setting

	Geometry
Classroom Teachers (Includes Special Population Educators)	21*
Higher Education	6*
Curriculum	1
School Administration	2
Special Education	2*

*Participants reported multiple assignments

Method

The Bookmark procedure was used to determine recommended cut scores for distinguishing performance on the Regents Examination in Geometry (Common Core). The Bookmark procedure is an *item-based* mapping method. It requires panelists to determine which items can be successfully answered two-thirds of the time by students at the boundaries between adjacent performance levels. The scaled difficulty value that separates the items that students at the threshold can answer two-thirds of the time from those they cannot answer is the cut score used to distinguish student performance into performance levels. The procedure typically involves three components: PLDs,

ordered item booklets (OIBs), and item maps. Each component is briefly described below.

Performance Level Descriptions (PLDs)

PLDs are the foundation of standard setting activities because they provide the explanation of how student performance differs from one performance level to the next (Perie, 2008). In fact, PLDs are of such influence that, in a well-run standard setting workshop, they determine the rigor of the performance and thus the decisions made about placement of the cut score (Perie, Hess, & Gong, 2008). Moreover, PLDs serve multiple purposes in terms of communicating policy, facilitating test development, guiding standard setting, and providing score interpretation. Three types of PLDs (Egan, Schneider & Ferrara, 2012) are used as an organizing framework for developing PLDs for the Regents Examinations:

- Policy PLD Policy Statements—Policy statements are designed to capture the vision that an agency has for its performance levels. They specify the number of levels and the names for each level and summarize the expectations of student performance for a testing program, including any policy decisions being made at particular levels.
- Range PLDs—Range PLDs are designed to describe the full range of performance for examinees at a given performance level. In other words, Range PLDs describe the aspects of test content or specific items that are indicative of a range of students at a specific performance level. Range PLDs can be informative in guiding item and test development as a testing program evolves. Range PLDs are also critical in that they are used to articulate a key component for standard setting, the Threshold PLDs. Note that the PLD meeting held in February was designed to produce Range PLDs.
- Threshold PLDs—Threshold PLDs (also known as Target PLDs) are designed to articulate the transition points between the different ranges of performance defined by the Range PLDs. Specifically, Threshold PLDs describe the knowledge and skills a student at the border between performance levels should know and be able to do. Because they articulate the specific performance that distinguishes levels of performance, Threshold PLDs are typically used in standard setting activities. Range PLDs and Threshold PLDs are clearly interdependent, which necessitates that they be developed in conjunction with each other.

Ultimately, PLDs are designed to describe the competencies of each performance level in relation to grade-level content standards while concurrently addressing their different functions. PLDs play a critical role in the standard setting process.

Ordered Item Booklet (OIB)

Within the Bookmark procedure, participants review the OIB, which is a booklet of the items from the operational test that have been ordered from easiest to hardest. Multiple-choice items appear along with their answer choices in the OIB, with each item printed on a single page. Constructed-response items appear along with their scoring rubrics multiple times because each item is worth multiple points. Specifically, each non-zero score point for a constructed-response item is presented in the OIB.

To sequence the items from easiest to hardest, a difficulty estimate for each item must be determined. Difficulty estimates supporting Bookmark standard setting are typically obtained using item response theory models that express item difficulty and student achievement on the same reporting scale. The Rasch measurement model (Rasch, 1960, 1980) was used to estimate item difficulty for selected-response items on the Regents Examinations. The Partial-Credit model (Andrich, 1978) was used to estimate item difficulty estimates for each score point for constructed- response models. These models are described in more detail below.

Rasch and Partial-Credit Models

The Rasch model applicable to dichotomously scored items (MC) can be expressed in the most familiar form of the model:

1. $$\Pr(correct \mid \beta_n, \delta_i) = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}.$$

The probability of success for a person with ability β_n on an item with difficulty δ_i is determined by the difference between the ability of the student and the difficulty of the item.

With the partial-credit model used for open-ended items, π_{nik} is the probability that person *n* will score *k* on item *i*. Then, the *first* threshold for item *i* is a score of 1 rather than a 0, which is the conditional probability of a score of 1, given a score of 0 or 1:

2. $$\Phi_{1ni} = \frac{\pi_{ni1}}{\pi_{ni0} + \pi_{ni1}} = \frac{\exp(\beta_n - \delta_{i1})}{1 + \exp(\beta_n - \delta_{i1})},$$

where β_n is the ability of person *n* and δ_{i1} is the difficulty of the first threshold. The expression on the right is identical to the Rasch model for a dichotomous item. The only differences are that now $\pi_{ni0} + \pi_{ni1} < 1$, since more than two response categories are provided, and δ_{i1}, while still the difficulty of the first threshold for item *i*, is not the difficulty of the only threshold for the item.

For example, with a three-point open-ended item, where a person *n* must achieve one of the four possible scores (0, 1, 2, or 3) on item *i*,

3. $$\pi_{ni0} + \pi_{ni1} + \pi_{ni2} + \pi_{ni3} = 1 \quad .$$

These relationships can be rearranged to obtain one general expression for the probability of person *n* scoring *x* on item *i*:

4. $$\pi_{nix} = \frac{\exp \sum_{j=1}^{x} (\beta_n - \delta_{ij})}{1 + \sum_{k=1}^{m_i} \exp \sum_{j=1}^{k} (\beta_n - \delta_{ij})}, \qquad x = 1,...,m_i \; .$$

If the number of thresholds (m_i) is one, the summations in expression (4.) drop out and it reduces to expression 1.

Using the operational response data from a representative sample of test takers, item difficulty parameter *b* was calibrated using WINSTEPS. Within the Rasch model, the item difficulty estimate produced by WINSTEPS assumes a 0.50 response probability. However, in standard setting, item difficulty estimates are typically computed relative to a response probability of two-thirds (i.e., 0.67). For dichotomous items, this required adding a factor of 0.69315 to the item difficulty parameters obtained from WINSTEPS to account for the increased response probability.

To obtain difficulty values for each score point within a constructed-response item using a two-thirds response probability, it was necessary to estimate the ability level associated with getting each score point or above. That is, for a four-point item, the ability associated with the likelihood of achieving two points or greater two-thirds of the time, three points or greater, and four points are estimated. This computation is done algorithmically, using a procedure detailed in Cizek and Bunch (2007).

After all difficulty estimates associated with a two-thirds response probability were computed, the OIB was created by ordering items in sequence of the difficulty estimates. Table 5 below includes information about the operational test and the OIB. Note that each page of the OIB included an annotation with the scaled difficulty estimate, key, and content standard.

Table 5. Composition of Ordered Item Booklet: Geometry

Part	Number of Items	Score Point Range	Number of OIB Pages
Part 1	24	0–1	24
Part 2	7	0–2	14
Part 3	3	0–4	12
Part 4	2	0–6	12
TOTAL	**36**	**-**	**62**

Item Map

The item map provides a corresponding document to the OIB. Essentially, the item map consists of information extracted from the OIB and presented in tabular form. The item map is presented with one row per item/point. The items/points are presented in difficulty sequence from easiest to hardest similar to the OIB. Each row includes the following information:

- Page number in OIB
- Original position on test form
- Content/standard identification
- Correct answer for selected-response items
- Score point and maximum score point for each constructed-response item
- Space for notes

Bookmark Judgment Task

During a standard setting using the Bookmark procedure, panelists review the test items ordered by difficulty from easiest to hardest. Item by item, panelists are asked to judge the likelihood that a student at the threshold between performance levels (e.g., the student who is just barely at Level 4) would answer the question correctly or achieve a particular score on a constructed-response item two-thirds of the time. The panelists are reminded throughout the process to use the policy guidance and the associated PLDs as the frame of reference. Panelists have typically been given an orienting task to become very familiar with the policy decisions and range PLDs in order to help articulate the knowledge and skills of students at the threshold. Panelists review the OIB information and make judgments for one PLD at a time in a specific sequence.

The specific judgment task with the Bookmark method requires panelists to evaluate whether students at the threshold of a PLD (e.g., just barely at level 4) have a chance of answering an item correctly or getting a particular score on a constructed-response item at a given response probability. The chance of answering (i.e., the response probability) that is typically used within Bookmark standard setting is two-thirds. Panelists are asked

to look at each item and evaluate whether a student at the threshold has at least a two-thirds chance of getting this item correct. For constructed-response items, the judgment task is to determine whether the student at the threshold has at least a two-thirds chance of achieving a certain number of points or higher on that item.

Panelists are instructed to move through the OIB, read each page/item in sequence, and evaluate the knowledge and skills as described by the PLDs that are required to respond to the item correctly (or to get the score point). Panelists are asked to identify the location in the ordered item booklet where the likelihood for a student at a given threshold to get an item right drops below the response probability of two-thirds. Panelists are asked to place a bookmark between the two items, marking the location where this transition occurs for this given threshold. Panelists then begin the process again for the next threshold until all thresholds have been bookmarked. This process is repeated over multiple rounds, with feedback after each round.

After each round, panelists have bookmarked pages that identify where in the OIB they feel each transition from one performance level to another is located. Given that each page within the OIB has an associated difficulty estimate expressed on a common metric, panelists have identified a cut score that can be used to distinguish student performance into two performance levels. Bookmark placements are translated back into the scale of measurement used to estimate item difficulties. The median of these difficulty estimates provided by the panelists is the recommended cut score for a given performance level.

Data

Data used to support these meetings were obtained from representative samples of students who had been administered the Regents Examinations immediately prior to the standard setting meetings. The samples were drawn to be representative of the typical population taking these Regents Examinations during a June administration. In order to expedite the production of the standard setting materials, a representative sample was selected in advance and processed ahead of remaining State materials. Item difficulty values, order item sequence, item maps, and impact data used at the standard setting meeting were all compiled using the data from this representative sample.

A preliminary sample was identified, using test enrollment data with a series of stratification values that included gender, ethnicity, English language learner (ELL) status, student with disabilities (SWD) status, socio-economic status, need/resource capacity (NRC) category, and previous performance on the applicable Regents Examination. Schools identified as being included in the sample received different answer documents for expedited processing by DRC. Some minor adjustments to the preliminary sample were made to account for differences between enrollment information and actual test administrations.

Summary statistics for the sample versus the population of a typical June administration (June 2014, in this case) are reported in Table 6. Note that the differences between the sample selected and the typical populations taking the Regents Examinations are negligible, suggesting that the information presented to standard setting panelists was well estimated.

Table 6. Sample vs. Population Summary, Geometry

		Population		Sample (9541 Students)		
		N	Pct.	N	Pct.	Pct. Diff
ETHNICITY	Asian	16490	10.37	1004	10.52	0.15
	Black	23004	14.47	1332	13.96	-0.51
	Hispanic	28884	18.16	1770	18.55	0.39
	American Indian / Native	682	0.43	30	0.31	-0.11
	Multiracial	1684	1.06	56	0.59	-0.47
	Pacific Islander	336	0.21	15	0.16	-0.05
	White	87942	55.3	5334	55.91	0.6
ENGLISH LANGUAGE LEARNER (ELL)	N	153866	96.76	9256	97.01	0.26
	Y	5156	3.24	285	2.99	-0.26
Need/Resource Capacity	High Need: New York City	45414	28.56	2819	29.55	0.99
	High Need: Large Cities	4985	3.13	298	3.12	-0.01
	High Need: Urban/Suburban	8116	5.1	423	4.43	-0.67
	High Need: Rural	7251	4.56	417	4.37	-0.19
	Average Need	48395	30.43	3002	31.46	1.03
	Low Need	27823	17.5	1711	17.93	0.44
	Charter School	2990	1.88	180	1.89	0.01
	Non-Public School	14048	8.83	691	7.24	-1.59
POVERTY	N	98497	61.94	5916	62.01	0.07
	Y	60525	38.06	3625	37.99	-0.07
GENDER	F	82699	52	5007	52.48	0.47
	M	76323	48	4534	47.52	-0.47
Student with Disabilities	N	149382	93.94	9008	94.41	0.48
	Y	9640	6.06	533	5.59	-0.48

Procedure

The standard setting was completed on June 16 and 17, 2015. The agenda for the standard setting meeting can be found in Appendix B.

Table Leader Training

Table leaders arrived the evening before the beginning of the standard setting meeting for training. Table-leader training consisted of an overview of the meeting agenda and the Bookmark procedure. Samples of materials provided for the standard setting were presented, and the role of table leaders was reviewed. Table leaders were to facilitate discussion and help participants stay focused at specific stages during the standard setting meeting.

Large Group Training

After the greetings and initial introductions, Senior Deputy Commissioner Wagner provided opening remarks and set the context for the meeting. A highlight of his presentation was an overview of the policy decisions associated with each performance level. These are shown below in Figure 1.

Figure 1. Policy Statements for Performance Levels

Performance Levels on Common Core Regents Exams
Level 5: Exceeds Common Core expectations
Level 4: Meets Common Core expectations (First required for Regents Diploma purposes with the Class of 2022)
Level 3: Partially meets Common Core expectations (Required for current Regents Diploma purposes. We expect comparable percentages of students to attain Level 3 or above as do students who pass current Regents Exams (2005 Standards) with a score of 65 or above)
Level 2 (Safety Net): Partially meets Common Core expectations (Required for Local Diploma purposes. We expect comparable percentages of students to attain Level 2 or above as do students who pass current Regents Exams (2005 Standards) with a score of 55 or above)
Level 1: Does not demonstrate Knowledge and Skills for Level 2

Following the Senior Deputy Commissioner's remarks, DRC provided an overview of the standard setting methodology. The major components of the Bookmark procedure were discussed in detail, including the PLDs and the OIB and its associated item map. Two procedures to be implemented within the Bookmark context were presented to the panelists.

1. Given the policy decision to hold the percentage of students at Level 3 and above as well as Level 2 and above to similar levels as those obtained in the previous Regents Examinations (see Figure 1), a policy validation exercise would be conducted. In particular, the bookmark locations that maintain consistency with the previous percentages would be pre-identified for panelists. Panelists would be asked to choose one of the pre-identified bookmark locations, consistent with the policy directive. Feedback on the bookmark placement would be gathered. This exercise would be completed as a single activity, and recommended cut scores for these two levels would be incorporated into the subsequent standard setting activity.

2. For the Level 4 and Level 5 cut scores, a traditional Bookmark standard setting procedure would be implemented. Results from the first activity would be incorporated so that panelists would see impact data for all performance levels.

There were 28 educator panelists for Geometry; each educator was pre-assigned to one of five tables. A table leader had also been pre-assigned to each table.

Following a break, panelists reviewed the test. The goals of the test review were for panelists to get a sense of the student experience in taking the Regents Examination and for panelists to preview the test items to be used in the standard setting. A subset of the items was identified for panelists to answer and score to ensure that the activity was not cursory. Panelists were instructed to review the remaining items.

Following the test review, DRC content facilitators led a discussion of the Level 4 cut score. This discussion asked for knowledge and skill statements describing students at the thresholds. Each table, working with several assigned domains of content, identified knowledge and skill statements that best described students at the thresholds. The synthesis of these statements across tables constituted the Threshold PLD and provided a frame of reference for the Bookmark task. Panelists, working in groups, repeated this process for Level 5.

Subject-specific training in the Bookmark standard setting method was then provided. The critical objective of the training was to ensure that the panelists understood the task being presented to them. Components of the training for panelists included a discussion of their role in the process, a detailed description of all steps in the Bookmark method, and practice exercises that contained publically available New York State assessment items. The point of the practice exercises was to provide hands-on experience with the steps and allow panelists to address additional questions that they might have once they had practiced. A copy of the training slides is provided in Appendix C. Once training was completed, a survey was taken to ensure that all panelists were ready to

proceed. All panelists indicated that they understood the task and were ready to proceed.

The policy verification task was first implemented for Level 2 and Level 3. For this task, the bookmark locations that resulted in equivalent passing rates relative to the previous Regent Examination were pre-identified. That is, a set of bookmark locations where the resulting percentage of students at Level 2 and above and Level 3 and above were identified. A color-coded item map provided the location of bookmark locations that would be consistent with the policy directives. Panelists were instructed to review the policy directives and the PLDs and identify which of the potential bookmark locations they would recommend. Panelists were reminded that the number of bookmark locations consistent with the State policy directives was relatively few and appeared early in the test book. Panelists completed a rating form to indicate their selected bookmark locations consistent with the State policy and completed a survey to demonstrate that they understood the policy verification task.

Once the policy verification was completed, the standard setting process for Level 4 and Level 5 was then started within each room. Three rounds were conducted. Each round is described below.

Round 1. Panelists were asked to identify the last item in the OIB that a threshold student at a given performance level would have a two-thirds chance of answering correctly. The bookmark location that panelists were to mark in the OIB was the last item that a student at the threshold could answer correctly two-thirds of the time; the student would not be expected to correctly answer the items that appeared later in the booklet. Panelists were asked to consider the knowledge and skills required to respond correctly to each progressively more difficult item. Panelists were reminded not to focus too much on a single item but to focus on the progression of items instead. Panelists were reminded that the OIBs were based on analysis of data selected from a representative sample from the June 2015 administration. It was emphasized that the work for this round was to be done individually.

Round 2. Table-level results from round 1 were provided to table leaders. Table-level results included the bookmark locations (i.e., pages selected by panelists) for each panelist and the median bookmark location for each performance level at the table. The panelists were asked to think about how similar their ratings were relative to the other panelists at their tables. Table leaders facilitated group discussion about differences/similarities, using the table-level results. Panelists were reminded that consensus was not a requirement and that differences should be discussed in order to provide additional insight into why such differences existed. After the group discussion, panelists were given the opportunity to revise their bookmark placements in a subsequent round.

Round 3. Table-level results from round 2 were provided to table leaders. Table-level results included the bookmark locations (i.e., pages selected by panelists) for each panelist and the median bookmark location for each performance level at the table. The panelists were asked to think about how similar their ratings were relative to the other panelists at their tables. Table leaders facilitated group discussion about differences/similarities, using the table-level results. Panelists were reminded that consensus was not a requirement and that differences should be discussed in order to provide additional insight into why such differences existed.

After table discussions were complete, the DRC facilitator presented table-level results as well as the room-level results to the full group. In particular, the median bookmark locations for all tables, as well as the room-level median of table-level medians, were presented. Panelists were then invited to discuss the table-level and room-level results, comparing and contrasting differences between tables and providing their initial feedback regarding the room-level results.

Once discussion of the table-level and room-level bookmark locations was complete, impact data based on the representative sample were provided to panelists. Specifically, the percentages of students at the different levels of performance were provided to panelists. After this discussion was complete, panelists were given another opportunity to revise their bookmarks.

After round 3 rating and analysis were completed. The final recommendations for bookmark locations, as well as the associated impact data, were presented. Panelists were also invited to provide any additional feedback about the PLD documents.

Results

Tables 7 through 9 provide summary information for all performance levels for Geometry across all three rounds of standard setting. The median bookmarked page for each table and the associated median difficulty estimate are provided. The difficulty estimate is based on a two-thirds response probability and is expressed on the logit scale used within the Rasch model. The room-level summary, computed as the median of table-level medians, is also presented. Because only one round was held for the policy verification of Levels 2 and 3, that information is repeated throughout the tables so that the results across all four levels can be compared.

Table 7. Median Bookmarked Pages, Geometry, Round 1

	Level 1/2		Level 2/3		Level 3/4		Level 4/5	
Table	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit
1	4	-0.7737	7	-0.1917	33.5	0.9184	52	1.5653
2	4	-0.7737	7	-0.1917	35	0.9535	49	1.4346
3	4.5	-0.7737	7	-0.1917	39	1.0309	55	1.7599
4	5	-0.4607	7	-0.1917	39	1.0309	48	1.3546
5	4	-0.7737	7	-0.1917	26	0.6069	48	1.3546
Room	4	**-0.7737**	**7**	**-0.1917**	**35**	**0.9535**	**49**	**1.4346**

Table 8. Median Bookmarked Pages, Geometry, Round 2

	Level 1/2		Level 2/3		Level 3/4		Level 4/5	
Table	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit
1	4	-0.7737	7	-0.1917	33	0.9184	48	1.3546
2	4	-0.7737	7	-0.1917	29	0.7393	53	1.6955
3	4.5	-0.7737	7	-0.1917	35	0.9535	55	1.7599
4	5	-0.4607	7	-0.1917	39	1.0309	49	1.4346
5	4	-0.7737	7	-0.1917	29	0.7393	48	1.3546
Room	4	**-0.7737**	**7**	**-0.1917**	**33**	**0.9184**	**49**	**1.4346**

Table 9. Median Bookmarked Pages, Geometry, Round 3

	Level 1/2		Level 2/3		Level 3/4		Level 4/5	
Table	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit	Median OIB Page	Logit
1	4	-0.7737	7	-0.1917	34	0.9409	48	1.3546
2	4	-0.7737	7	-0.1917	32	0.8883	50	1.4553
3	4.5	-0.7737	7	-0.1917	35	0.9535	54	1.7252
4	5	-0.4607	7	-0.1917	38	0.9991	49	1.4346
5	4	-0.7737	7	-0.1917	29	0.7393	48	1.3546
Room	4	**-0.7737**	**7**	**-0.1917**	**34**	**0.9409**	**49**	**1.4346**

Comparisons between rounds also indicate that the cut score recommendation did not fluctuate much. Impact data were presented at the beginning of round 3. The additional information had a negligible effect on the subsequent recommendations that the group made in round 3.

Figures 2 and 3 represent the percentage of students in each performance level, using the cut score recommendation after rounds 2 and 3 for Geometry. The impact data were based on a representative sample of students who were administered the 2015 Regents Examination. Note that these were the figures that were used to present impact data to panelists.

Figure 2. Percentage of Students in Performance Levels, Geometry, Round 2.

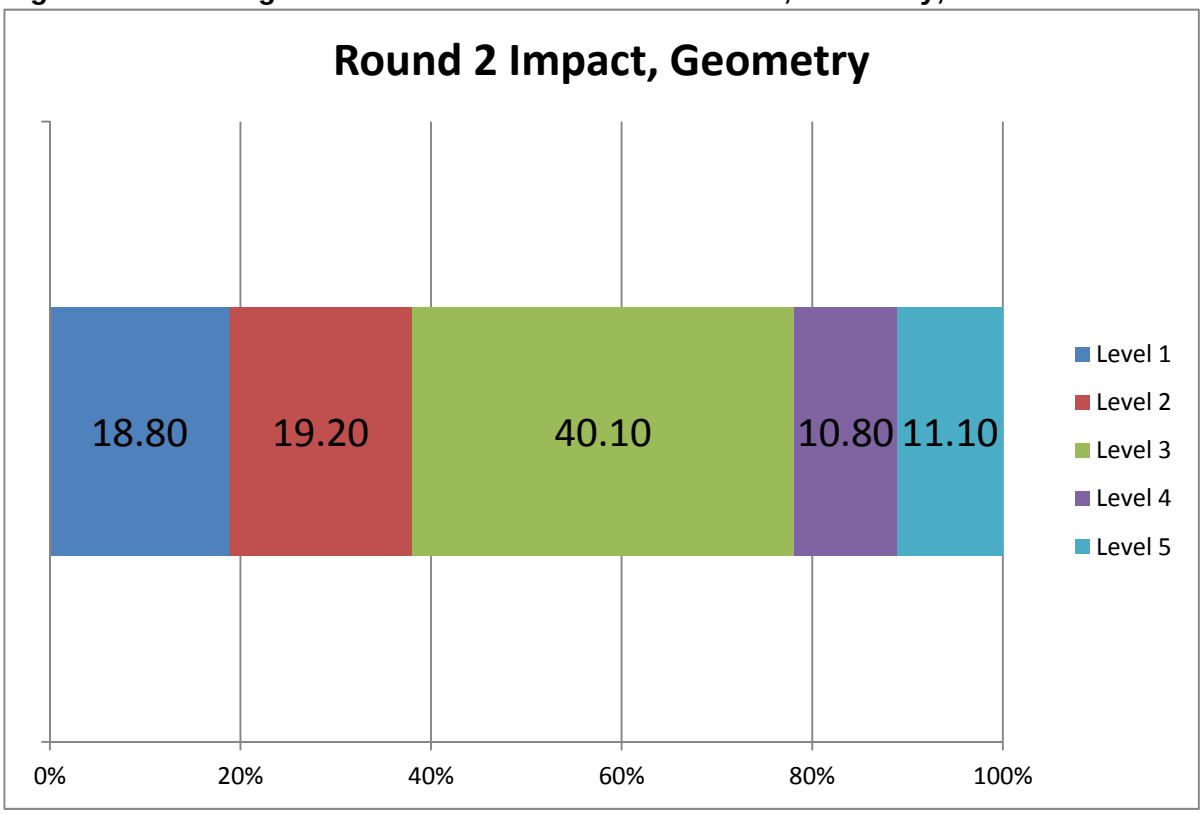
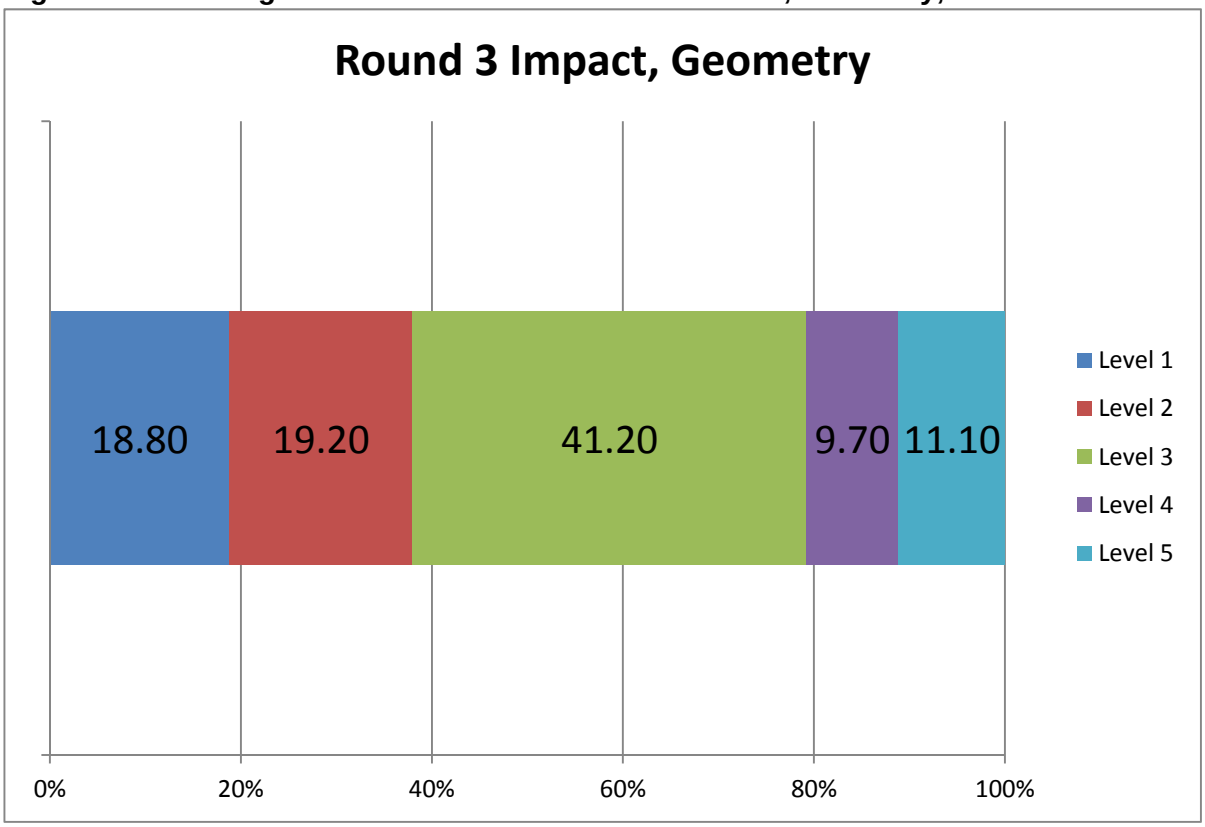


Figure 3. Percentage of Students in Performance Levels, Geometry, Round 3



.

Evaluations

An exit survey was completed by each panelist after the policy verification of the Level 2 and Level 3 cut scores. Panelists answered the survey questions using a scale of 1–4, with 1 being "strongly disagree" and 4 being "strongly agree." The survey questions and the results for each question are provided in Appendix D.

The intentions of this exit survey were to ensure that panelists understood the policy directives to place constraints on the overall standard setting process and to get their feedback about the recommended cut scores, given the policy directives. Over 95% of the panelists moderately or strongly agreed that they understood the policy directives and that the projected bookmarks fairly represented the minimal level of achievement for students at Level 2 and Level 3.

An additional exit survey was completed by each panelist after all standard setting activities were completed. Panelists answered the survey questions using a scale of 1–4, with 1 being "strongly disagree" and 4 being "strongly agree." The survey questions and the results for each question are provided in Appendix E.

The intentions of this exit survey were to gather feedback on different aspects of the standard setting procedure and to get panelists' feedback on the recommended cut scores and associated results. All of the panelists moderately or strongly agreed that the cut scores accurately represented the PLDs. All of the panelists felt that the Bookmark standard setting method and associated activities would produce appropriate results for New York State students.

Final Recommendations

As described in the previous sections, the NYSED, with facilitation by DRC, conducted a formal standard setting that consisted of two meetings. The first meeting was devoted to the development of PLDs that articulate the range of knowledge, skills, and proficiencies of students at the five levels of performance specified by State policy. The second meeting was dedicated to the identification of cut scores consistent with the PLDs and State policy directives, using a standardized, scientific procedure called the Bookmark method.

Both meetings reflected best psychometric practices as articulated in the Standards for Educational and Psychological Measurement, and proceeded according to the plans reviewed by the New York State Technical Advisory Committee. The participants in both meetings were diverse and representative of the State. All groups followed, without incident, instructions delivered by standard setting staff. All activities were formally overseen by the Office of State Assessment senior management and psychometric staff.

After careful consideration of the nature of the new examinations, the rigor of the new curricula, the transitional and aspirational aspects of the State policy directives, and the role of the assessment in student learning throughout high school and beyond, the standard setting committees made recommendations on the cut scores to the Commissioner of Education. The Commissioner accepted the recommendations of the standard setting panelists. The approved cut scores were provided to the NYSED's scaling and equating contractor for implementation within the scale of measurement used to report student performance on the New York State Regents Examinations.

The standard setting process was developed and implemented with great care, and best practices in assessment and psychometrics were followed. The policy decisions implemented were consistent with sound psychometric research to guarantee an effective and efficient standard setting.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–73.

Cizek, G. J., & Bunch, M. B. (2007). Standard setting: A guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage.

Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. J. Cizek (Ed.), Setting performance standards: Foundations, methods, and innovations (2nd ed., pp. 79–106). New York: Routledge.

Lewis D. M., Mitzel, H. C., & Green, D. R. (1996). Standard Setting: A Bookmark Approach. In D. R. Green (Chair), IRT-Based Standard-Setting Procedures Utilizing Behavioral Anchoring. Symposium presented at the 1996 Council of Chief State School Officers 1996 National Conference on Large Scale Assessment, Phoenix, AZ.

Perie, M. (2008). A guide to understanding and developing performance-level descriptors. Educational Measurement: Issues and Practice, 27(4), 15–29.

Perie, M., Hess, K., & Gong, B. (2008). Writing performance level descriptors: Applying lessons learned from the general assessment to alternate assessments based on alternate and modified achievement standards. Paper presented at the Annual Meeting of the National Council on Measurement in Education in New York, NY, March 2008 and available at http://www.nciea.org.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press.

Appendix A: Range Performance Level Descriptions, Geometry

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Congruence (G-CO)	**Use precise language** to describe a sequence of rigid motions to determine the congruency of figures.	**Describe** a sequence of rigid motions **to determine the congruency of figures**.	Identify and draw a **sequence** of rigid motions in the plane **to verify the congruency of figures**.	Identify and draw a rigid motion in the plane.	Sketch triangles and rectangles.
	Use precise language to predict the effect of a given rigid motion on a given figure.	**Predict the effect** of a given rigid motion on a given figure.	Identify the image and **describe the effect** of a given rigid motion.	Identify the image of a given rigid motion.	
	Formulate a complete line of geometric reasoning to prove a geometric **theorem**.	**Formulate a complete line** of geometric reasoning to prove a specific geometric statement.	**Formulate a partial line** of geometric reasoning in an effort **to prove a specific geometric statement**.	Provide a correct geometric statement pertaining to the given geometric information.	Restate given information in the context of a proof.
	Use the rotations and reflections that carry a figure onto itself to prove or explain if the figure is or is not regular.	**Describe** the rotations and reflections that carry a figure onto itself.	**Identify** the rotations and reflections that carry a figure onto itself.		
	Determine the validity of geometric arguments and **revise invalid geometric arguments**.	Determine the validity of geometric arguments **with justification.**	Determine the validity of geometric arguments.		

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
(G-CO continued)	Make **advanced formal geometric constructions** using appropriate tools.	**Construct the application of the listed constructions,** for example, using the construction of a midpoint to construct the median of a triangle or construct the dilation of a figure not on the coordinate plane.	Make basic formal geometric constructions using appropriate tools. Examples of basic constructions include but are not limited to: copy a segment, bisecting a segment, bisecting an angle.	Construct rays, triangles, and angles.	Construct lines and line segments.

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Similarity, Right Triangles, and Trigonometry (G-SRT)	**Use precise language** to describe a sequence of similarity transformations to determine the similarity of figures.	**Describe** a sequence of similarity transformations **to determine the similarity of figures**.	**Identify a sequence of similarity transformations** in the plane **to verify the similarity of figures**.	Perform a dilation in the coordinate plane centered at the origin. Distinguish between a dilation and a translation, reflection, or rotation.	
	Formulate a complete line of geometric reasoning to prove a geometric **theorem**.	**Formulate a complete line** of geometric reasoning to prove a specific geometric statement.	**Formulate a partial line** of geometric reasoning in an effort **to prove a specific geometric statement**.	Provide correct geometric statements pertaining to the given geometric information.	Restate given information in the context of a proof.

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
(G-SRT continued)	Apply congruence or similarity criteria to solve **complex problems involving multiple concepts**, and explain the geometric reasoning involved.	Apply congruence or similarity criteria to solve problems, **and explain the geometric reasoning involved**.	Apply congruence or similarity criteria to **solve problems**.	Apply congruence or similarity criteria to solve simple problems.	
		Use the Pythagorean Theorem, trigonometric ratios, and the relationship between sine and cosine of complementary angles **to solve complex problems.**	**Use the Pythagorean Theorem, trigonometric ratios, and the relationship between sine and cosine of complementary angles to solve problems.**	Identify the trigonometric ratios of a right triangle.	Sketch and label the sides of right triangles.
	Determine the validity of geometric arguments and **revise invalid geometric arguments**.	Determine the validity of geometric arguments **with justification.**	Determine the validity of geometric arguments.		

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Circles (G-C)	Use appropriate tools to construct the inscribed and circumscribed circle for a given triangle **and justify the construction**.	Use appropriate tools to construct the **inscribed** and circumscribed circle for a given triangle.	Use appropriate tools to construct the **circumscribed** circle for a given triangle.		
	Derive the formula for the arc length and area of a sector.	**Apply formulas** for arc length and area of a sector **to solve complex problems**.	Determine the arc length and area of a sector given any central angle in degrees or **radians.**	Determine the area of a quarter, half, or three-quarter circle, given the area of the entire circle.	Write an expression for the area of a circle given the radius.
		Apply theorems about arcs, angles, and **segments** related to circles.	**Apply theorems** about arcs and angles related to circles.	Identify arcs, angles, and segments related to circles.	Visually compare central angle measures.
		Explain the radian measure of a central angle as the constant of proportionality between the arc length and the radius of a circle.	**Identify** central angles in different circles that have the same radian measure.		

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
(G-C continued)		**Formulate a complete line** of geometric reasoning to prove properties of angles for a quadrilateral inscribed in a circle.	Formulate a **partia**l line of geometric reasoning in an effort to prove properties of angles for a quadrilateral inscribed in a circle.	**Identify** a missing angle in a diagram involving a quadrilateral inscribed in a circle.	
		Formulate a complete line of geometric reasoning to prove that circles are similar.	Formulate a **partial** line of geometric reasoning in an effort to prove that circles are similar.	Find missing radius and circumference measurements using circle similarity.	

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Expressing Geometric Properties with Equations (G-GPE)	**Use the Pythagorean theorem to derive the equation of a circle**.	Given the equation of a circle in standard form, **complete the square to obtain the center and radius**.	**Identify the center and radius of a circle** when given the equation in center-radius form.	Complete the square with a single variable.	
	Use coordinates to formulate a complete line of geometric reasoning to **prove or disprove** a geometric theorem.	Use coordinates to **formulate a complete line** of geometric reasoning to prove a specific geometric statement.	Use numerical coordinates to formulate a partial line of geometric reasoning in an effort to prove a specific geometric statement.	Given three coordinates of a special quadrilateral, determine the fourth coordinate.	
	Explain why parallel lines have the same slopes and perpendicular lines have negative reciprocal slopes.	Use the slope criteria for parallel and perpendicular lines **to solve geometric problems**.	Identify the **equations of lines as parallel, perpendicular, or neither**.	Identify the slope of a line given its equation.	Distinguish between lines in a coordinate plane with positive and negative slopes.
		Identify **the rational** coordinates of a point that divides a segment into a given ratio.	Identify **the whole number** coordinates of a point **that divides a segment into a given ratio**.	Identify the coordinates of the midpoint of a line segment.	Locate the midpoint of a horizontal or vertical line in a coordinate plane.

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
(G-GPE continued)	Use coordinates to compute perimeters and areas of **compound figures**.	Use coordinates to compute perimeters of polygons and areas of triangles and rectangles with rational or **irrational bases and heights**.	Use coordinates to **compute perimeters of polygons with rational side lengths**. Use coordinates to **compute areas of triangles and rectangles with rational bases and heights**.	Compute the length of vertical, horizontal, and diagonal segments on the coordinate plane with integer coordinates. Compute the perimeter of polygons with integer side lengths in the coordinate plane. Compute the area of triangles and rectangles with integer bases and heights in the coordinate plane.	Identify the whole number coordinates of triangles and rectangles.

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Geometric Measurement & Dimensions (G-GMD)	Write a **formal argument** for the formulas for the circumference of a circle, area of a circle, and volumes of a cylinder, pyramid, and cone.	Write an **informal argument** for the formulas for the circumference of a circle, area of a circle, and volumes of a cylinder, pyramid, and cone.			
	Use the volume formulas for cylinders, pyramids, cones, and spheres to solve **modeling problems involving compound figures**.	Use the volume formulas for cylinders, pyramids, cones, and spheres **to solve modeling problems**.	Use the volume formulas for cylinders, **pyramids**, cones, and spheres **to find various dimensions of the solid**, such as finding the radius of a sphere given the volume.	Compute the volumes for cylinders, cones, and spheres.	Compute the volume of a rectangular prism with integer dimensions.

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
(G-GMD continued)	**Describe the similarities and differences between various cross-sections** of three-dimensional objects, such as explaining the difference between the areas of different cross-sections of the same figure.	**Describe** the two-dimensional cross-sections of three-dimensional objects.	Identify the two-dimensional cross-sections using a **diagram of a three-dimensional object**.	Identify a two-dimensional cross-section that results from slicing a right rectangular prism or a right rectangular pyramid.	Identify the shape of the base of a rectangular prism, triangular prism, or cylinder.
	Describe the similarities and differences between various rotations of two-dimensional objects, such as a half rotation or rotating about different axes.	**Describe** three-dimensional objects generated by rotations of two-dimensional objects.	**Identify** three-dimensional objects generated by rotations of two-dimensional objects.		

Domain	NYS Level 5	NYS Level 4	NYS Level 3	NYS Level 2	NYS Level 1
Modeling with Geometry (G-MG)	**Create a model to solve real-world problems**, which may include applying density to real-world situations or solving design problems.	Apply geometric concepts in modeling situations to **solve complex real-world problems**, which may include applying density to real-world situations or **solving design problems**.	Apply **concepts of density to solve a problem** that may include **converting between two- and three-dimensional units.**	**Given two of the three values in the density formula, find the third value.**	Compute the area of a rectangular region, given whole number dimensions.

Appendix B: Agenda for Standard Setting

NEW YORK STATE REGENTS EXAMINATIONS IN GEOMETRY (COMMON CORE)
STANDARD SETTING
JUNE 16-17, 2015

AGENDA

Tuesday, June 16, 2015

8:00 am – 8:30 am	**Registration and Breakfast**
8:30 am – 10:15 am	**Welcome and Training**
	The purpose of the day's first session is to provide background information on this standard setting meeting and articulate your roles and responsibilities in the standard setting process. A detailed overview of the process being used will be given.
10:15 am – 10:30 am	**Break**
10:30 am – 11:45 pm	**Introductions and Test Review**
	The goal of the test review is to review the operational test individually to get a sense of the student experience and to preview the test items that will be used in the bookmark process.
11:45 pm – 12:45 pm	**Lunch**
12:45 pm – 2:45 pm	**Review PLDs and Discuss Threshold Students**
	The goal of this discussion is to develop a common understanding of the students at each threshold and to articulate a description of students at the thresholds.
2:45 pm – 3:00 pm	**Break**
3:00 pm – 3:30 pm	**Refresher Training**
	The purpose of this training session is to re-orient you to the bookmark method and go through a practice activity.
3:30 pm – 4:00 pm	**Level 3 and Level 2 Bookmark Evaluation**
	The purpose of this session is to evaluate the policy directive in relation to the Level 3/Level 2 and Level 2/Level 1 cut scores.
4:00 pm – 5:00 pm	**Level 4 and Level 5 Bookmarking - Round 1**
	During round 1, you will individually determine the bookmark placement for the thresholds based on the threshold PLDs and your professional expertise. These bookmark placements will be translated to cut scores for the exam.

(over)

Wednesday, June 17, 2015

8:00 am – 8:30 am	**Breakfast**
8:30 am – 10:00am	**Discussion of Round 1 Results** The goal of this session is to discuss and gain perspective of table peers regarding round 1 bookmark placements. A consensus does not need to be reached.
10:00 am – 10:15 am	**Break**
10:15 am – 11:30 am	**Level 4 and Level 5 Bookmarking - Round 2** During round 2, you will individually determine the bookmark placement for the thresholds based on the threshold PLDs and your professional expertise. These bookmark placements will be translated to cut scores for the exam.
11:30 am – 12:30 pm	**Lunch**
12:30 pm – 2:00 pm	**Discussion of Round 2 Results** The goal of this session is to discuss and gain perspective of all subject peers regarding round 2 bookmark placements. The room facilitator will share overall recommended bookmark cut scores as well as impact data based on the cut scores.
2:00 pm – 3:30 pm	**Level 4 and Level 5 Bookmarking - Round 3** During round 3, you will individually determine the bookmark placement for the thresholds based on the threshold PLDs and your professional expertise. These bookmark placements will be translated to cut scores for the exam.
3:30 pm – 4:00 pm	**Break**
4:00 pm – 4:30 pm	**Discussion of Round 3 Results** Final impact results based on the recommended cut scores will be shared and reactions to the bookmark process and impact results will be discussed.

Appendix C: Training Slides

engage^ny
Our Students. Their Moment.

New York State Regents Examination in Geometry (Common Core) Standard Setting

Albany, New York
June 16-17, 2015

Welcome and Introductions

- Data Recognition Corporation
  - Marc Julian, Trainer and Lead Facilitator
  - Pete Tressel, Project Manager
  - Ben Sorenson, Data Analyst
  - John Selisky, Geometry Content Specialist

Opening Session Agenda

- Welcome
- NYSED Remarks
- Purpose of Meeting
- Standard Setting Overview
- Methodology
- Administrative Issues

Welcome and Introductions

New York State Department of Education
- Ken Wagner, Senior Deputy Commissioner
- Candy Shyer, Assistant Commissioner
- Zach Warner, Project Coordinator
- Paul Anderson, Geometry Lead

What is Standard Setting?

- A process that lets experts make judgments about the content that students at each level of performance should know

Performance Levels

Performance Levels on Common Core Regents Exams
Level 5: Exceeds Common Core expectations
Level 4: Meets Common Core expectations (First required for Regents Diploma purposes with the Class of 2022)
Level 3: Partially meets Common Core expectations (Required for current Regents Diploma purposes. We expect comparable percentages of students to attain Level 3 or above as do students who pass current Regents Exams (2005 Standards) with a score of 65 or above)
Level 2 (Safety Net): Partially meets Common Core expectations (Required for Local Diploma purposes. We expect comparable percentages of students to attain Level 2 or above as do students who pass current Regents Exams (2005 Standards) with a score of 55 or above)
Level 1: Does not demonstrate Knowledge and Skills for Level 2

engage^ny
Our Students. Their Moment.

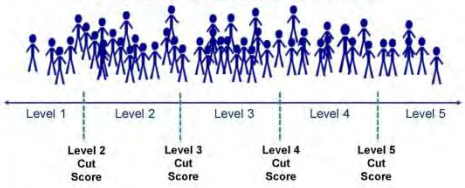
New York State Regents Examination in Geometry (Common Core) Standard Setting

Albany, New York
June 16-17, 2015

Why do Standard Setting?

- We need to decide what score on the Regents exams tells us whether a student's performance was enough to be considered "Level 4"

How are Standards Set?



Phases of Standard Setting

Policy Statements

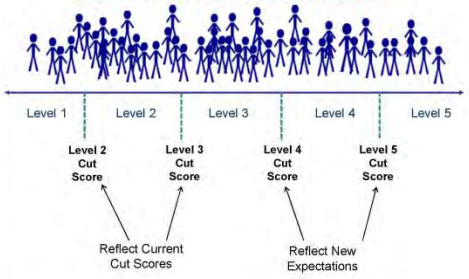
- Broad statements that define level of rigor and/or policy implication for each performance level

Level 4: Meets Common Core expectations (First required for Regents Diploma purposes with the Class of 2022)
Level 3: Partially meets Common Core expectations (Required for current Regents Diploma purposes. We expect comparable percentages of students to attain Level 3 or above as do students who pass current Regents Exams (2005 Standards) with a score of 65 or above)

Performance Levels

Performance Levels on Common Core Regents Exams
Level 5: Exceeds Common Core expectations
Level 4: Meets Common Core expectations (First required for Regents Diploma purposes with the Class of 2022)
Level 3: Partially meets Common Core expectations (Required for current Regents Diploma purposes. We expect comparable percentages of students to attain Level 3 or above as do students who pass current Regents Exams (2005 Standards) with a score of 65 or above)
Level 2 (Safety Net): Partially meets Common Core expectations (Required for Local Diploma purposes. We expect comparable percentages of students to attain Level 2 or above as do students who pass current Regents Exams (2005 Standards) with a score of 55 or above)
Level 1: Does not demonstrate Knowledge and Skills for Level 2

How are Standards Set?



Phases of Standard Setting

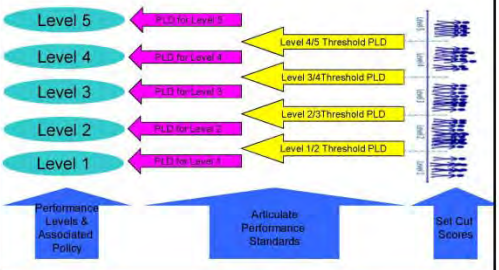
Performance Level Descriptions

- Performance Level Descriptions:
  - Describe the knowledge and skills expected of students at each of the five performance levels
  - Describe the range of performance of the level, not the boundary between levels
  - Developed in previous meetings by NYS educators

Conceptualize the Threshold Student

- Based on the PLDs, visualize New York State students who are:
  - Just barely entering the next higher level
  - What knowledge and skills should a student have at the thresholds?

Phases of Standard Setting



Standard Setting Method

- Bookmark Procedure
  - Focuses on items
  - Focuses on students at the threshold between performance levels
  - Focuses on what a student should know and be able to do, as defined by the performance level descriptions (PLDs)

Overview of Method

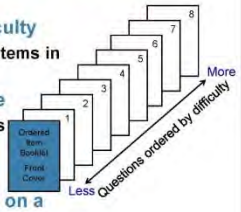
- Take the test
- Review performance level descriptions
- Develop and review threshold students descriptions
- Multiple rounds of bookmark placements
  - Work individually
  - Group discussion during 2nd and 3rd
  - Feedback after each round

Take the test

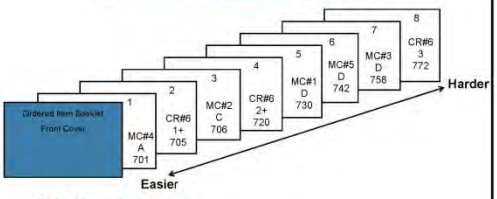
- Objectives
  - Take a sample of test items
  - Provide you with insight into the student testing experience
  - Think about how students would have experienced the test
  - What knowledge and skills does each item require?

Ordered Item Booklet

- Items are ordered by difficulty
  - Easy items in front; hard items in back
- One MC/CR point per page
  - CRs appear multiple times
  - Scoring rubric for each CR score point
- Difficulty estimates based on a sample of NYS students that is representative of a typical Regents Exam administration

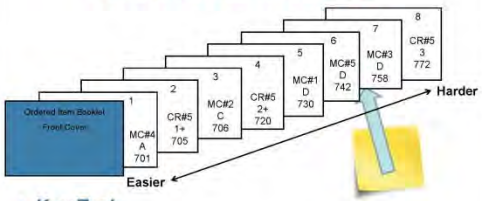


Ordered Item Booklet



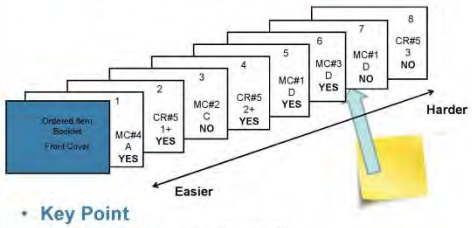
- OIB for 6 item test
  - 5 MC items, 1 CR item worth 3 points
  - 8 Pages in the OIB
  - Answer key, scaled difficulty value, and CR rubric

Ordered Item Booklet



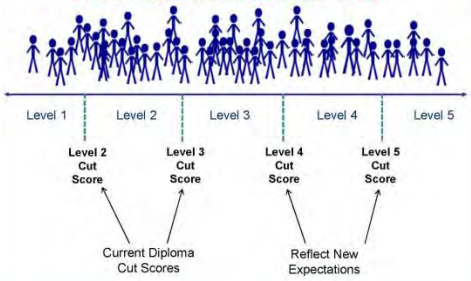
- Key Task
  - Place a bookmark between items that mark the transition between one level to another
  - Do that for each threshold

Ordered Item Booklet



- Key Point
  - Do not focus on a single item
  - Identify groups of items where the transition occurs

How are Standards Set?



Outcomes

- This committee is making a set of recommendations
- Cut scores for these performance levels will be recommended to the Acting Commissioner of Education, subject to the approval of the Board of Regents

Administrative Issues

- Roles and Responsibilities
  - NYSED Representatives
  - Project Managers
  - Room Facilitators
  - Content Specialists
  - Table Leaders
  - Panelists
- ELA Focus Group Meeting Tomorrow

Administrative Issues

- Security
  - Non-disclosure agreement
  - All materials stay in room
  - Notify the room facilitator for unscheduled breaks
  - Discussion outside of room
    - Do NOT discuss:
      - Results, PLDs, Items
    - DO discuss process

Administrative Issues

- Other meeting guidelines
  - No cell phones, tablets, or laptops at tables
  - Sit quietly if others are still working
  - No materials leave the room
  - Start on time
  - Participate!

Forms and Documents

- Non-Disclosure Agreement
- Travel Expense Reimbursement Form
- Substitute Teacher Reimbursement Form
- Panelist Survey

Next Steps

- Complete forms and give to DRC staff:
  - Non-disclosure agreement
  - Panelist survey
- Take a 15 minute break and return here

Questions?

engage^ny
Our Students. Their Moment.

New York State Regents Examination in Geometry (Common Core) Standard Setting

Albany, New York
June 16-17, 2015

Today's Agenda

- Introductions
- Take the Test
- Lunch
- Review PLDs and Create Threshold Descriptions
- Training
- Level 3 and Level 2 Review
- Level 4 and Level 5 Bookmarking - Round 1

Introductions

- What is your name?
- Where are you from?
- What is your current role in education?

Take the Test

- Please take the following items on the test:
  - MC – 4, 7, 8, 11, 14, 15, 17, 20, 22, 24
  - CR – 27, 28, 29, 30, 34, 35, 36
- Please review remaining items if you have time
- When done, please sign your test book in
- Report back to this room after lunch

LUNCH

- Served in Great American Grill
- Meet back here at 12:45pm

Threshold Student Descriptions

- For each threshold:
  - Review performance level descriptions
  - Generate knowledge and skill statements
  - Develop summary of knowledge and skill statements for threshold students
  - In the following sequence
    - Level 3/4 threshold
    - Level 4/5 threshold

Review the PLDs

- Performance Level Descriptions:
  - Describe the knowledge and skills expected of students at each of the five performance levels
  - Describe the range of performance of the level, not the boundary between levels

Conceptualize the Threshold Student

- Based on the PLDs, visualize New York State students who are:
  - Just barely entering the next higher level
  - What knowledge and skills should a student have at the thresholds?
  - We are focusing on the thresholds between performance levels that reflect new expectations

Conceptualize the Threshold Student

Level 4: Meets Common Core expectations
(First required for Regents Diploma purposes with the Class of 2022)

- Divide the Geometry domains among the tables
- At each table, develop knowledge and skill statements for each threshold
  - Operationalize each statement for the threshold students – i.e. what would be "just enough" or "just barely" sufficient for each PLD statement
- Consider behaviors and classroom experiences directly linked to the PLDs
- Focus on knowledge and skills
  - Avoid other students attributes (e.g., low SES)
- Form a group definition (concept) of a threshold student discussing across tables

Conceptualize the Threshold Student

Level 5: Exceeds Common Core expectations

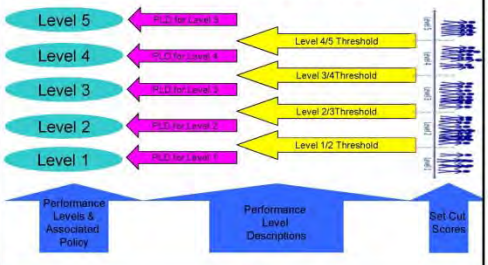
- Divide the Geometry domains among the tables
- At each table, develop knowledge and skill statements for each threshold
  - Operationalize each statement for the threshold students – i.e. what would be "just enough" or "just barely" sufficient for each PLD statement
- Consider behaviors and classroom experiences directly linked to the PLDs
- Focus on knowledge and skills
  - Avoid other students attributes (e.g., low SES)
- Form a group definition (concept) of a threshold student discussing across tables

Break

- Return to this room at 3pm

Bookmark Training and Practice

Relationship of PLDs, Performance Levels and Cut Scores

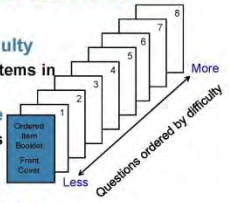


Materials

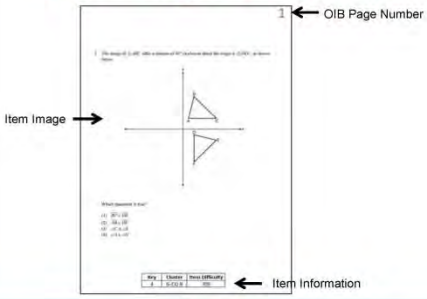
- PLDs
- Description of Threshold Students
- OIB
- Item Map
- Item Separation Chart
- Bookmark Form

Ordered Item Booklet

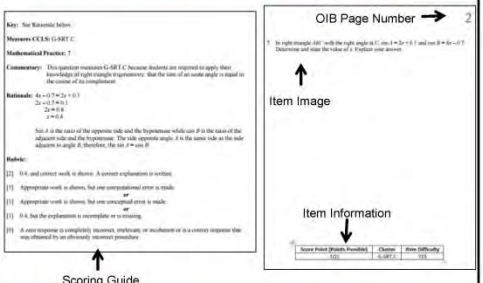
- Items are ordered by difficulty
  - Easy items in front; hard items in back
- One MC/CR point per page
  - CRs appear multiple times
  - Scoring rubric for each CR score point
- Difficulty estimates based on a sample of NYS students that is representative of a typical Regents Exam administration



OIB - Multiple-Choice Item



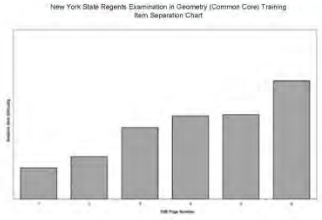
OIB – Constructed-Response Item



Item Map

OIB Page	Test Seq	Score Pt (Pts Poss)	Key	Cluster	Item Difficulty	Notes
1	2		4	G-GMD.B	701	
2	4	1 (2)		G-SRT.C	719	
3	1		1	G-SRT.C	726	
4	5		4	G-GPE.B	754	
5	6	2(2)		G-SRT.C	765	
6	3		3	G-C.B	859	

Item Separation Chart

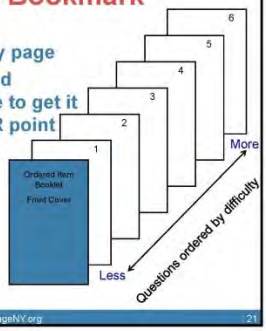


Bookmark Form

	OIB Page Number Your "bookmark" will be between 2 pages – write down the first page. (This will be the last item the threshold student should be able to answer correctly 2/3 of the time.)
	Level 3/Level 4
Round 1	
Round 2	
Round 3	

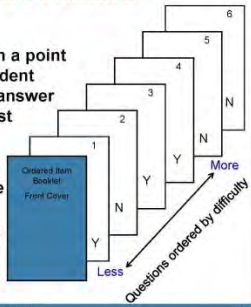
Placing a Bookmark

- Go through OIB page by page
- Judge whether threshold students should be able to get it correct or score that CR point or higher
  - Use 2/3rds as criterion



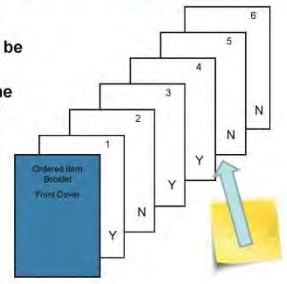
Placing a Bookmark

- Key Task:
  - Continue until you reach a point where the threshold student would consistently not answer the item correctly at least 2/3rds of the time
- Key Points
  - Do not focus on a single item
  - Identify groups of items where the transition occurs



Placing a Bookmark

- Place Bookmark:
  - Your bookmark will be between 2 pages
  - Place a post-it on the last item where you think the threshold student should get the item correct 2/3rds of the time
  - Record this page number on your bookmark form



Important Points

- Regents Exams are constructed using the statistical model called *item response theory (IRT)*. All state testing programs and most certification tests use IRT
- IRT allows different test forms with different items to have scale scores with the same meaning within the same subject area (e.g., June and August Geometry Regents Exams)
- With IRT, scaled scores are designed to award points for not just how many items a student gets correct, but the challenge-level of the item
- Our raw-to-scale conversion charts reflect this and the bookmark standard setting method is designed with IRT

Important Points

- The ordered item booklets are based on IRT
- Each item is a representation of the knowledge and skills required to achieve a particular scaled score
- **The page number is not the same as the raw score (number correct/points achieved)**
- Thinking about bookmark pages as if they were raw scores is inconsistent with the bookmark methodology

Example

HYPOTHETICAL

OIB Page	Raw Score
1	15
2	21
3	24
4	26
5	27
6	28

Remember

- Do not consider the page number as a proxy for number of points achieved on the test!

Practice Exercise

- **Review the sample Geometry items**
- **Using the threshold descriptions that we created, visualize a student just barely out of the Level 3 and just barely into Level 4**
- **Go through the training OIB page by page**
  - **assess whether a just barely Level 4 student has a sufficient probability of answering each item correctly**
    - Sufficient is defined as 2/3rds of the time

Practice Exercise

- **For each item, indicate on the item map or the OIB if you expect the threshold student to answer the item correctly at least 2/3rds of the time (Y) or less than 2/3rds (N)**
- **Place a post-it note on the last item you judge that your threshold student would get correct *at least* 2/3rds of the time**
- **Indicate on the training bookmark placement form the last item you judge your threshold student would get correct *at least* 2/3rds of the time**

Practice

Practice Exercise: Sample Results

- After round 1:
  - Individual cut scores (i.e., OIB page) given to each table leader
  - Table cut score (median) provided to each table
- After round 2:
  - Individual and table cut scores provided to each table
  - Table cut scores and overall cut score recommendation shown to entire room

	Round 1			
	Table 1 OIB Pages			
Panelist ID	I/II	II/III	III/IV	IV/V
1			5	
2			2	
3			2	
4			3	
5			5	
6			3	
Median			3	

	Round 1 OIB Page Cut			
Table	I/II	II/III	III/IV	IV/V
1			3	
2			3	
3			3	
4			4	
5			4	
6			2	
7			3	
Room			3	

Practice Exercise: Sample Results

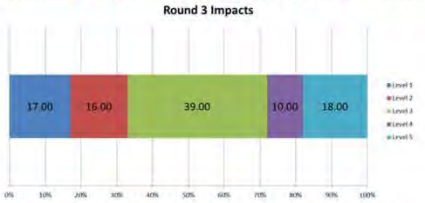
Round 1 Impacts				
Level 1	Level 2	Level 3	Level 4	Level 5
18.00	22.00	35.00	11.00	14.00

Round 2 Impacts				
Level 1	Level 2	Level 3	Level 4	Level 5
16.00	20.00	32.00	10.00	22.00

Round 3 Impacts				
Level 1	Level 2	Level 3	Level 4	Level 5
17.00	16.00	39.00	10.00	18.00

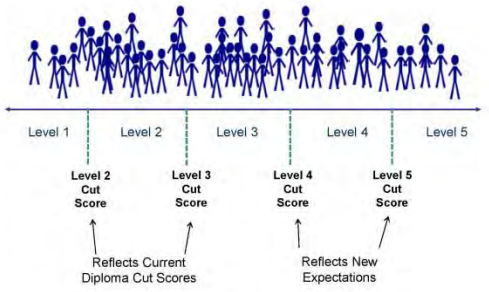
- After rounds 2 and 3, impact data across rounds will be presented

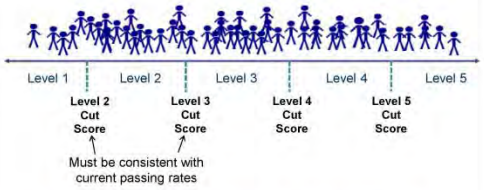
Practice Exercise: Sample Results



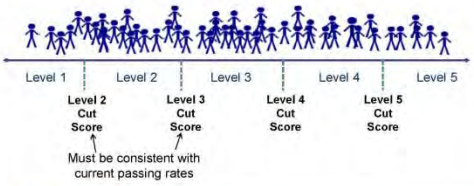
- After rounds 2 and 3, data will be provided showing the percent of students that would be in each performance level based on the recommended cut scores

Questions?

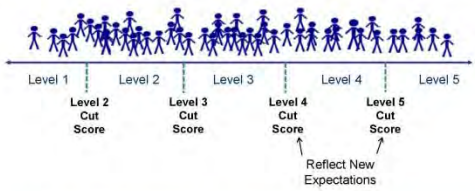




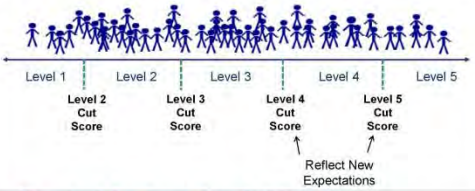
- We provide the location of the bookmarks in the OIB that are consistent with this policy
- We provide a narrow range of items where the bookmarks can be moved that maintain the policy



- **Your task**
  - Review narrow range of items around the bookmark
  - Decide whether to move the bookmark within the range based on your expert judgment and PLD discussions
  - Does not require multiple rounds



- **Policy Decisions for upper 2 cut scores**
  - Level 4 and Level 5 cut scores reflect new expectations for students
  - By 2022, the Level 4 cut score will be required for a Regents diploma



- **Your task for these thresholds is to**
  - Review each item in sequence and ask yourself whether a student at a given threshold would get an item right most of the time (2/3rds)
  - Identify where in the OIB the answer to that question transitions from Yes to No
  - Multiple rounds with feedback after each round

Questions?

- **Complete Readiness Form**

Policy Review: Level 3 and Level 2

Policy Review: Level 3 and Level 2

- **Policy directive requires that the percent of students at or above these two levels should be consistent with current levels**
- **Passing rates were reviewed from prior administrations**

Policy Review: Level 3 and Level 2

- Looking at performance on the new Regents Exam, we worked backwards and identified:
  - Bookmark pages consistent with policy directive
  - Narrow range of bookmark locations
- When you apply passing rates from older test to new more rigorous test, the bookmark placements will appear early in the OIB
  - Remember bookmark location is not the same as number of points
  - These bookmark locations translate to reasonable number correct scores

Policy Review: Level 3 and Level 2

- Provide you with small item map that includes
  - Range of bookmark locations
    - Blue for Level 3
    - Green for Level 2
  - % of students at or above each page

Policy Review: Level 3 and Level 2

- Task
  - What page number that is in line with the policy directive would you recommend be used?
  - Please provide your rationale
    - Knowledge and skills reflected in items
    - Discussion of PLDs
    - Expert judgment
    - Impact data

Bookmarking Activities Level 4 and Level 5

Round 1

Start Round 1

- Ask yourself the following questions for each threshold
  - MC Items: Should a student just barely at a threshold be able to get this right 2/3 of the time?
  - CR Items: Should a student just barely at a threshold be able to get at least this score point 2/3 of the time?
- Remember:
  - Threshold Student Descriptions
  - Following order:
    - Level 3/Level 4
    - Level 4/Level 5
  - Individual task

Next Steps

- Sign in your OIB when you are done
- Breakfast will be available starting at 7:30 tomorrow morning in the Fitzroy Room
- Meet in this room by 8:30 tomorrow morning

Day 2 Agenda

- Discuss Round 1 Results
- Break
- Round 2
- Lunch
- Discuss Round 2 Results
- Round 3
- Break
- Discuss Round 3 Results

Discuss Round 1 Results

- Table leader will lead table-level discussions for each threshold:
  - What is the distribution of bookmark pages?
  - How did you determine your bookmark placement?
    - Use threshold PLD summaries to defend your placement

Break

- Meet back here at 10:15

Start Round 2

- Ask yourself:
  - MC Items: Should a just barely Level 4 student be able to get this right 2/3 of the time?
  - CR Items: Should a just barely Level 4 student be able to get at least this score point 2/3 of the time?
- Remember:
  - Threshold Student Descriptions
  - Following order:
    - Level 3/ Level 4
    - Level 4/ Level 5
  - Individual task

Lunch

- Served in Fitzroy
- Meet back here at 12:30

Discuss Round 2 Results

- Table leader will lead table-level discussions for each threshold:
  - Did the distribution of bookmark pages change?
  - How did you determine your bookmark placement?
    - Use threshold PLD summaries to defend your placement
- Cross table discussion for each threshold:
  - Are there notable differences across tables
- Impact data

Start Round 3

- Ask yourself:
  - MC Items: Should a just barely Level 4 student be able to get this right 2/3 of the time?
  - CR Items: Should a just barely Level 4 student be able to get at least this score point 2/3 of the time?
- Remember:
  - Use PLDs
  - Following order:
    - Level 3 / Level 4
    - Level 4 / Level 5
  - Individual task
- When complete, sign in your materials and complete the evaluation form.

Thank you!

Appendix D: Policy Verification for Level 2 and Level 3 Bookmark Placements, Exit Survey and Results, Geometry

Geometry (Common Core) Level 2/Level 3 and Level 1/Level 2 Exit Survey Results

1. I understand the Board of Regents policy directive to place constraints on the overall standard setting process, such that the percentage of students who score at Levels 2 and 3 and above on the Common Core Regents Exams will remain comparable to those percentages of students who scored at a 55 and 65 and above on the current Regents Exams (2005 Standards).

Valid *N*	Percent Selecting Category				Avg.[1]	*SD*
	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree		
27	59.26	37.04	3.70	0.00	3.56	.58

[1]Strongly Agree = 4, Moderately Agree = 3, Moderately Disagree = 2, Strongly Disagree =1

2. The impact data (percentages of students at or above the suggested cut scores) presented were helpful to me in evaluating the cut scores.

Valid *N*	Percent Selecting Category				Avg.[1]	*SD*
	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree		
25	24.00	56.00	20.00	0.00	3.04	.68

[1]Strongly Agree = 4, Moderately Agree = 3, Moderately Disagree = 2, Strongly Disagree =1

3. I believe that my Level 2/Level 3 cut score fairly represents the minimal level of achievement for students at Level 3, given the policy directive.

3 Valid *N*	Percent Selecting Category				Avg.[1]	*SD*
	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree		
28	42.86	50.00	3.57	3.57	3.29	.57

[1]Strongly Agree = 4, Moderately Agree = 3, Moderately Disagree = 2, Strongly Disagree =1

4. If you answered Moderately Disagree or Strongly Disagree to Question 3, indicate whether you believe the cut score is too high or too low and provide your rationale.

	Percent Selecting Category	
Valid *N*	Too High	Too Low
2	50.00	50.00

Rationale:

No response

5. I believe that my Level 2/Level 1 cut score fairly represents the minimal level of achievement for students at Level 2, given the policy directive.

5	Percent Selecting Category					
Valid *N*	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	Avg.[1]	*SD*
28	35.71	50.00	10.71	3.57	3.18	.77

[1]Strongly Agree = 4, Moderately Agree = 3, Moderately Disagree = 2, Strongly Disagree =1

6. If you answered Moderately Disagree or Strongly Disagree to Question 5, indicate whether you believe the cut score is too high or too low and provide your rationale.

	Percent Selecting Category	
Valid *N*	Too High	Too Low
4	100.00	0.00

Rationale:

NYS Level 2 for congruence says “Identify the image of a given rigid motion”.

I believe page numbers 4 and 5 reflected the same level (Level 2) so it was difficult to set question 4 in Level 1.

Students without the skills to proceed are entering then failing upper math. I have students in Geo who can’t do Alg – but passed w/30 cut score – only knowing 30/86 of Algebra.

Feel that students could not successful pass item #3 at the Level 1.

Geometry (Common Core) Exit Survey and Results

1. Please rate the extent of your agreement with each statement regarding the opening session:

	Valid *N*	Percent Selecting Category					
		Strongly Agree	Agree	Disagree	Strongly Disagree	Avg.[1]	SD
The opening session provided a clear description of the meeting's goals.	28	71.42	28.58	0.00	0.00	3.71	.46
The opening session helped me understand my tasks.	28	53.57	42.86	3.57	0.00	3.50	.58
The opening session leaders clearly explained the procedures.	28	57.14	42.86	0.00	0.00	3.57	.50
The opening session addressed many of my questions and concerns.	28	42.86	57.14	0.00	0.00	3.43	.50

[1]Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree =1

2. Please rate the extent of your agreement with each statement regarding the Geometry training session:

	Valid *N*	Percent Selecting Category					
		Strongly Agree	Agree	Disagree	Strongly Disagree	Avg.[1]	SD
The training session leader clearly explained the procedures.	28	64.29	35.71	0.00	0.00	3.64	.49
The training session leader clearly explained the materials used in the bookmark process.	28	71.42	28.58	0.00	0.00	3.71	.46
The training helped me understand my tasks.	28	57.14	42.86	0.00	0.00	3.57	.50
The training addressed many of my questions and concerns.	28	42.86	57.14	0.00	0.00	3.43	.50
The training materials were effective in preparing for subsequent tasks.	28	53.57	46.43	0.00	0.00	3.54	.50
The practice exercises were useful.	27	44.44	40.74	14.81	0.00	3.30	.72

[1]Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree =1

3. Please rate the extent of your agreement with each statement regarding the performance level descriptions (PLDs):

	Valid *N*	Percent Selecting Category					
		Strongly Agree	Agree	Disagree	Strongly Disagree	Avg.[1]	SD
Adequate information was provided to panelists regarding the PLDs.	27	40.74	59.26	0.00	0.00	3.41	.50
Adequate time was provided for panelists to gain understanding of the PLDs.	27	55.55	44.44	0.00	0.00	3.56	.51
The PLDs communicate a reasonable profile of students' achievement at each level.	27	44.44	51.85	3.70	0.00	3.41	.57

[1]Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree =1

4. Please rate the extent of your agreement with each statement regarding the threshold performance level descriptions (PLDs):

	Valid *N*	Percent Selecting Category					
		Strongly Agree	Agree	Disagree	Strongly Disagree	Avg.[1]	SD
Adequate time was provided for panelists to articulate the threshold PLDs.	27	59.26	40.74	0.00	0.00	3.59	.50
The threshold PLDs communicate a reasonable profile of students' achievement at each threshold.	27	48.15	48.15	3.70	0.00	3.44	.58

[1]Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree =1

5. Please indicate your opinion regarding the usefulness of the following materials used:

	Valid *N*	Percent Selecting Category				Avg.[1]	SD
		Very Useful	Useful	Somewhat Useful	Not Useful		
Performance level descriptions	28	60.71	39.29	0.00	0.00	3.61	.50
Operational test book	28	53.57	35.71	10.71	0.00	3.43	.69
Ordered item booklet	28	75.00	25.00	0.00	0.00	3.75	.44
Item map	28	32.14	46.43	14.29	7.14	3.04	.88
Item separation chart	28	28.57	60.71	10.71	0.00	3.18	.61
Statistical impact data	28	42.85	39.29	14.29	3.57	3.21	.83

[1]Very Useful = 4, Useful = 3, Somewhat Useful = 2, Not Useful =1

6. Please indicate the extent of your satisfaction with the following roles:

	Valid *N*	Percent Selecting Category				Avg.[1]	SD
		Very Satisfied	Satisfied	Partially Satisfied	Not Satisfied		
DRC psychometric lead	28	85.71	14.29	0.00	0.00	3.86	.36
DRC room facilitator	28	78.57	21.43	0.00	0.00	3.79	.42
DRC content specialist	28	78.57	21.43	0.00	0.00	3.79	.42
Other DRC Staff	27	85.19	14.81	0.00	0.00	3.85	.36

[1]Very Satisfied = 4, Satisfied = 3, Partially Satisfied = 2, Not Satisfied =1

7. Please indicate your opinion regarding the amount of time allotted for each activity:

	Valid *N*	Percent Selecting Category				
		Too Little Time	About Right	Too Much Time	Avg.[1]	SD
Training	28	3.57	67.86	28.57	2.25	.52
PLD discussion	28	3.57	71.43	25.00	2.21	.50
Round 1 ratings	28	10.71	75.00	14.29	2.04	.51
Round 1 discussion	28	0.00	82.14	17.86	2.18	.39
Round 2 ratings	28	0.00	78.57	21.43	2.21	.42
Round 2 discussion	28	10.71	71.43	17.86	2.07	.54
Round 3 ratings	28	0.00	78.57	21.43	2.21	.42

[1]Too Little Time = 1, About Right = 2, Too Much Time = 3

8. Please indicate the level of confidence you had in placing the bookmark location for each assessment cut score:

	Valid *N*	Percent Selecting Category					
		Very Confident	Confident	Partially Confident	Not Confident	Avg.[1]	SD
Level 3/Level 4 cut score	28	60.71	35.71	3.57	0.00	3.57	.57
Level 4/Level 5 cut score	28	60.71	35.71	3.57	0.00	3.57	.57

[1]Very Confident = 4, Confident = 3, Partially Confident = 2, Not Confident =1

9. Please rate the extent of your agreement with each statement regarding the processes and results:

	Valid *N*	Percent Selecting Category					
		Strongly Agree	Agree	Disagree	Strongly Disagree	Avg.[1]	SD
The processes and methods used will produce appropriate results.	25	40.00	60.00	0.00	0.00	3.40	.50
My bookmark placements accurately represent the PLDs.	25	60.00	40.00	0.00	0.00	3.60	.50

[1]Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree =1

Page intentionally left blank