

New York State Regents Examination in Algebra II (Common Core)

Standard Setting Technical Report



Prepared for the New York State Education Department
by Pearson

2016

Copyright

Developed and published under contract with the New York State Education Department by Pearson.

Copyright © 2016 by the New York State Education Department.

Secure Materials.

All rights reserved. No part of this document may be reproduced or transmitted by any means. Use of these materials is expressly limited to the New York State Education Department.

Table of Contents

Executive Summary	1
Regents Examination in Algebra II (Common Core)	2
Performance Level Descriptions (PLDs)	3
Performance Level Description Workshop	4
Test Sample	5
Standard Setting	6
Panelists	6
Methodology	7
Pre-workshop	7
Workshop	8
Test Review	9
PLDs	9
Modified Angoff Judgment Training	9
Standard Setting Rounds	10
Level 2 Task Force	12
Cut Scores and Impact Data	12
Workshop Evaluation	14
Final Recommendations	15
References	16
Appendix A: Standard Setting Workshop Agenda	17
Appendix B: Panelist Readiness	19
Appendix C: Workshop Evaluation	21
Table 1. Test Blueprint for Regents Examination in Algebra II (Common Core)	2
Table 2. Regents Examination in Algebra II (Common Core) Design	3
Table 3. New York State Regents Examination Policy PLDs	4
Table 4. Geographic Locations of Panelists for PLD Workshop	4
Table 5. 2016 Test Sample for Standard Setting	6
Table 6. Geographic Locations of Panelists for Standard Setting Workshop	6
Table 7. Standard Setting Pre-workshop PLD Assignments	8
Table 8. Table Assignment of Algebra II (Common Core) Domains	9
Table 9. Mock Table of Items with Greatest Variability in Judgments	11
Table 10. Recommended Cut Scores Across Rounds	13

Figure 1. Round 1 Impact Data	13
Figure 2. Round 2 Impact Data	14
Figure 3. Round 3 Impact Data	14

Executive Summary

Standard setting was conducted for the New York State Regents Examination in Algebra II (Common Core) and consisted of two parts: a performance level description (PLD) workshop and a standard setting workshop for recommending cut scores. The primary goal for these workshops was to recommend performance standards, or 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.

The PLD workshop was held on February 29, 2016 in Albany, New York. The focus of this meeting was the review and revision of draft PLDs, which describe the specific knowledge and skills of students at each level of performance. Each PLD is designed to describe the range of students at that performance level and is used in the subsequent standard setting meeting.

The standard setting meeting was held June 15–16, 2016, in Albany, New York. The purpose of this meeting was to recommend four cut scores that distinguish the five levels of performance on the New York State Regents Examination in Algebra II (Common Core). Using the PLDs articulated in February, panelists followed the Modified Angoff standard setting procedure, resulting in cut score recommendations that were brought to the New York State Education Department.

In this report, panelists, materials, methodologies, and results are presented for the New York State Regents Examination in Algebra II (Common Core) standard setting.

Regents Examination in Algebra II (Common Core)

In June 2016, the New York State Regents Examination in Algebra II (Common Core) measuring the NYS P-12 Common Core Learning Standards (CCLS) was administered for the first time. The questions on the Regents Examination in Algebra II (Common Core) will measure the Common Core Algebra II standards as specified in the Partnership for Assessment of Readiness for College and Careers (PARCC) Model Content Framework for Algebra II. Algebra II is associated with high school content standards within five conceptual categories: Number & Quantity, Algebra, Functions, Geometry, and Statistics & Probability. Table 1 shows the percent of test credit and the domains included in Algebra II for each conceptual category.

Table 1. Test Blueprint for Regents Examination in Algebra II (Common Core)

Conceptual Category	Percent of Test By Credit	Domains in Algebra II
Number & Quantity	5-12%	The Real Number System (N-RN) Quantities (N-Q) The Complex Number System (N-CN)
Algebra	35-44%	Seeing Structure in Expressions (A-SSE) Arithmetic with Polynomials and Rational Expressions (A-APR) Creating Equations (A-CED) Reasoning with Equations and Inequalities (A-REI) Expressing Geometric Properties with Equations (G-GPE) ¹
Functions	30-40%	Interpreting Functions (F-IF) Building Functions (F-BF) Linear, Quadratic, and Exponential Models (F-LE) Trigonometric Functions (F-TF)
Statistics & Probability	14-21%	Interpreting Categorical and Quantitative Data (S-ID) Making Inferences and Justifying Conclusions (S-IC) Conditional Probability and the Rules of Probability (S-CP)

The Regents Examination in Algebra II (Common Core) contains four parts: one with multiple-choice (MC) questions and three with constructed-response (CR) questions. For MC questions, students select the correct response from four answer choices. Some MC items require completion of multiple steps. For CR questions, students are required to clearly indicate the necessary steps, including appropriate formula substitutions,

¹ Although the organization of the CCLS places one standard from the G-GPE domain into the Geometry Conceptual Category, the content within this domain will be assessed as part of the Algebra Conceptual Category for the Regents Examination in Algebra II (Common Core).

diagrams, graphs, proofs, etc. In some cases, they may be required to provide written explanations or justifications to demonstrate conceptual understanding. For the CR items, there are 2-, 4-, and 6-credit questions. Table 2 shows the test design for the Regents Examination in Algebra II (Common Core).

Table 2. Regents Examination in Algebra II (Common Core) Design

Test Component	Number of Questions	Credits per Question	Total Credits in Section
Part I	24	2	48
Part II	8	2	16
Part III	4	4	16
Part IV	1	6	6
Total	37	-	86

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, and 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, and Ferrara, 2012) are used as an organizing framework for developing PLDs for the Regents exams:

- **Policy PLD statements**—Policy PLD 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. Note that Table 3 provides the Policy PLDs for the New York State Regents Examination in Algebra II (Common Core).
- **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 a PLD workshop was held in February 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.

Table 3. New York State Regents Examination Policy PLDs

Level 5: Exceeds expectations of the standards
Level 4: Meets expectations of the standards
Level 3: Partially meets expectations of the standards (sufficient for Regents Diploma purposes)
Level 2 (Safety Net): Partially meets expectations of the standards (sufficient for Local Diploma purposes)
Level 1: Does not demonstrate the knowledge and skills for Level 2

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.

Performance Level Description Workshop

Range PLDs for the Regents Examination in Algebra II (Common Core) were drafted under the supervision of the New York State Education Department (NYSED). These descriptions outline the knowledge and skills expected of students at each of New York State's five performance levels. Once drafted, the Range PLDs were presented to a panel of New York State educators for review. This workshop took place on February 29, 2016, in Albany. The purpose of this workshop was to review and revise the Range PLDs so that they can be used for standard setting, the process of recommending cut scores that differentiate test scores into the performance levels. Sixteen educators, recruited by NYSED, participated in the performance level workshop. There were nine females and seven males within this panel. Of the 16 panelists, six were from higher education. Table 4 provides a geographic summary of the panelists.

Table 4. Geographic Locations of Panelists for PLD Workshop

Geographic Location	Number of Panelists
Capital District	4
Central NY	2
Long Island	3
NYC	3
Western NY	2
Other	2

The workshop began with a welcome and introductions from NYSED staff. A Pearson facilitator provided an explanation of performance level descriptions, highlighting the distinction between the policy and range descriptions. After an explanation of range descriptions, the process of standard setting was introduced. Standard setting was briefly discussed to highlight the use of performance level descriptions in setting performance

standards. From this, the concept of Threshold PLDs was introduced. Threshold descriptions, the minimal knowledge and skills needed for entry into a performance level, was briefly discussed to highlight another layer of performance descriptions within the overall standard setting process. *Note: Panelists were encouraged to draft threshold descriptions as a secondary task during this workshop.*

After explanations of the performance level descriptions and the standard setting process, the workshop task was described. Panelists were instructed to review the Range PLDs and revise them as they felt necessary. For example, review the knowledge and skills for Level 3 in a given domain and determine if the identified knowledge and skills represent what a student should know and be able to do for graduation. In addition, the panelists were to determine if student performance is articulated across the performance levels and whether the verbs used are appropriate and observable. As mentioned before, panelists were asked to consider Threshold PLDs, but only after the primary task was completed.

Suggested revisions from the panelists were provided to NYSED for consideration. NYSED used the panelists' feedback to revise the Range PLDs for use during standard setting. In addition, some Threshold PLDs were submitted for standard setting.

Test Sample

A sample of schools was identified to obtain test score data for setting performance standards on the Regents Examination in Algebra II (Common Core). Schools were selected to form a representative sample in terms of demographic and achievement characteristics from data on the Regents Examination in Algebra 2/Trigonometry (2005 Standard) administered in 2015. After the 2016 administration of the Regents Examination in Algebra II (Common Core), additional sampling adjustments were made to obtain a representative sample for standard setting purposes. The final sample consisted of approximately 12,000 students. Table 5 contains a summary of the standard setting sample and the 2015 population characteristics for comparison.

Table 5. 2016 Test Sample for Standard Setting

		Sample		Population	
		N	%	N	%
All Students		11,727	100.00	93,435	100.00
Gender	Female	6,130	52.27	50,114	53.64
	Male	5,597	47.73	43,321	46.36
Ethnicity	Asian	2,198	18.74	13,486	14.43
	American Indian or Alaska Native	352	3	384	0.41
	Black or African American	1,302	11.1	10,938	11.71
	Hispanic or Latino	1,308	11.15	14,506	15.53
	Multiracial	116	0.99	1,090	1.17
	Native Hawaiian or Pacific Islander	68	0.58	212	0.23
	White	6,383	54.43	52,819	56.53
Students with Disabilities	No	11,426	97.43	90,893	97.28
	Yes	301	2.57	2,542	2.72
English Language Learner	No	11,633	99.2	91,766	98.21
	Yes	94	0.8	1,669	1.79
Economically Disadvantaged	No	7,429	63.35	60,904	65.18
	Yes	4,298	36.65	32,531	34.82
Need/Resource Capacity Category	High Need: New York City	4,271	36.42	29,717	31.80
	High Need: Large Cities	213	1.82	1,390	1.49
	High Need: Urban/Suburban	216	1.84	3,915	4.19
	High Need: Rural	459	3.91	4,123	4.41
	Average Need	4,121	35.14	29,127	31.17
	Low Need	2,312	19.72	16,910	18.10
	Charter School	78	0.67	1,386	1.48
	Nonpublic School	57	0.49	6,867	7.35

Standard Setting

Cut scores for the Regents Examination in Algebra II (Common Core) were recommended by a panel of 20 New York State educators over a two-day standard setting workshop. The Modified Angoff procedure (Angoff, 1971) of determining cut scores was used in a multi-round process of performance judgments, feedback data, and discussions.

Panelists

The panelists, recruited by NYSED, represented the major geographic regions of New York State, as shown in Table 6. There were 20 panelists that participated in standard setting. Eleven panelists were female and 9 were male. Of the 20 panelists, 5 were from higher education whereas the remaining 15 panelists served in positions in New York high schools.

Table 6. Geographic Locations of Panelists for Standard Setting Workshop

Geographic Location	Number of Panelists
Capital District	2
Central NY	1
Long Island	5
Lower Hudson	3
NYC	3
Western NY	6

From these panelists, two small groups were identified: table leaders and the Level 2 Task Force. Table leaders were chosen to primarily help facilitate small group discussions during the workshops. The Level 2 Task Force was chosen to recommend cut scores for Level 2 separate from the full panel of panelists, given the characteristics of the student subgroup that mostly represents this level of performance. This Task Force and its task will be discussed later in this report.

The evening prior to the standard setting workshop, the group of table leaders met with the workshop facilitator and NYSED to discuss workshop roles and responsibilities and an overview of the workshop. The facilitator provided a high-level overview of the workshop activities and outlined the role of the table leader within those tasks. Also, the facilitator introduced the standard setting procedure, the judgment task, and the methods by which panelists' judgments would be captured.

Methodology

The Modified Angoff standard setting procedure was used for this standard setting workshop. Panelists provided estimates of student success to each item of the Regents Examination in Algebra II (Common Core) for each performance level cut. For each test item, panelists provided an estimate of how well students at a given performance level should do. For MC items, panelists provided percentages of students at a given performance level that should answer each item correctly. For the CR items, panelists provided the item's score point that students at a given performance level should obtain.

The standard setting process, though, focuses on students *just barely* at the performance levels, or *threshold* students. Therefore, the judgments provided by the panelists for each item and performance level were in terms of success of threshold students. For example, *what percentage of students who are just barely at a given performance level should answer this item correctly or how many points should students just barely at a given performance level obtain.*

Pre-workshop

To engage in the judgment process of standard setting, there must be an understanding of content expectations for each performance level. Prior to the standard setting workshop, panelists were provided some pre-workshop tasks through the Moodle online platform: review of sample Algebra II (Common Core) items, review of Policy and Range PLDs, and development of Threshold PLDs. These tasks were provided ahead of the workshop in order to set the context for standard setting. Panelists were asked to review some test items—items available to the public as practice items—to understand some of what students had to do on the test.

In addition, the Policy and Range PLDs were provided so that panelists could review and understand the expectations within Algebra II (Common Core) content domains across state performance levels. For this task, panelists were assigned to particular content domains such that each panelist reviewed at least one domain and no panelist reviewed more than two domains. The general assignment by domain is provided in Table 7.

Table 7. Standard Setting Pre-workshop PLD Assignments

Domain	Number of Panelists	Comment
The Real Number System (N-RN)	1	N/A
Quantities (N-Q)	1	This panelist will also review N-CN.
The Complex Number System (N-CN)	1	This panelist will also review N-Q.
Seeing Structure in Expressions (A-SSE)	1	This panelist will also review A-CED.
Arithmetic with Polynomials & Rational Expressions (A-APR)	1	N/A
Creating Equations (A-CED)	1	This panelist will also review A-SSE.
Reasoning with Equations & Inequalities (A-REI)	2	N/A
Interpreting Functions (F-IF)	2	N/A
Building Functions (F-BF)	2	N/A
Linear, Quadratic, & Exponential Functions (F-LE)	2	N/A
Trigonometric Functions (F-TF)	2	N/A
Expressing Geometric Properties with Equations (G-GPE)	1	N/A
Interpreting Categorical & Quantitative Data (S-ID)	1	N/A
Making Inferences & Justifying Conclusions (S-IC)	2	N/A
Conditional Probability & the Rules of Probability (S-CP)	2	N/A

Panelists were assigned to specific domains so that they could develop Threshold PLDs to the assigned domain. Specific instructions were provided on how Threshold PLDs can be derived from the Range PLDs, using some of the Threshold PLDs that were provided during the PLD workshop. For developing Threshold PLDs, panelists were asked to review the assigned Range PLDs and provide one or two statements of knowledge and skills that classify students *just barely* in Level 3, Level 4, and Level 5. Panelists entered responses for each performance level directly into the Moodle platform. These responses were collected and presented during the standard setting workshop.

Workshop

The standard setting workshop was held in Albany, New York, June 15–16, 2016. Appendix A contains the workshop agenda. The workshop began with a welcome from NYSED, introductory remarks about the Regents Examination program, and the goals for setting performance standards on Algebra II (Common Core). A Pearson facilitator provided an overview of the standard setting process, explaining the different types of contextual information used (i.e., performance level descriptions, test content), the standard setting judgment process, and the different types of feedback data that would be presented throughout the workshop. After the general orientation, including workshop logistics, the panelists began the workshop by reviewing the operational Regents Examination in Algebra II (Common Core).

Test Review

The panelists were provided sets of MC and CR items to review. Accompanying the test booklet was a rating guide that provided the content standard assessed for each item and scoring rubric for each CR item. Similar to the goal of item review for the pre-workshop task, the panelists were to understand what students had to do on the test. A short discussion followed this task.

PLDs

After the test review, the facilitator discussed PLDs and their use during the standard setting process. An explanation of deriving Threshold PLDs from the Range PLDs was also presented. The Threshold PLDs provided by the panelists prior to the workshop were distributed for review and discussion. Each table of panelists was assigned to a set of Algebra II (Common Core) domains for review and discussion (see Table 8).

Table 8. Table Assignment of Algebra II (Common Core) Domains

Table	Domains
1	A-SSE, A-APR, A-CED, A-REI, and G-GPE
2	F-IF, F-BF, F-LE, and F-TF
3	N-RN, N-Q, and N-CN
4	S-ID, S-IC, and S-CP

Working with the PLDs occurred over two steps. The first step was for panelists to review the assigned Range PLDs and discuss key differences between the performance levels. For the second step, the panelists were instructed to review the Threshold PLDs provided for the assigned domains.² In reviewing the Threshold PLDs, the panelists were to understand the knowledge and skills for students just barely into the performance level. Panelists discussed the Threshold PLDs within the table groups and were instructed to take notes during their discussion. After panelists discussed the Threshold PLDs within their table, the facilitator reorganized the panelists into temporary small groups such that each group had a representative from the original domain groupings. The result was that each panelist could explain the discussions they had on their assigned Threshold PLDs (refer to Table 8) to those panelists who had reviewed other domains. Once this process was complete, all panelists had gained knowledge on Threshold PLDs for all Algebra II (Common Core) domains.

Modified Angoff Judgment Training

The facilitator discussed the judgment task for the Modified Angoff standard setting procedure, providing a demonstration of making item-level judgments for each performance level. This process includes an item-level judgment for each item for each performance level. As previously mentioned, the judgment task varied slightly by item type. For MC items, the panelists were to provide a percentage of students who are just barely at a given performance level that should answer the item correctly. For CR items, the panelists were provide the rubric score that students just barely at a given

² For two domains, Threshold PLDs were not provided during the pre-workshop activity and the impacted tables needed to create some descriptions as part of the discussion.

performance level should obtain. These judgments were to be made using the Threshold PLDs discussed earlier in the process. The entire judgment process involved three rounds, with each judgment round followed by feedback data (i.e., judgment summaries) and discussions. Multiple rounds are often used in the standard setting process to allow panelists to revise their judgments after peer discussions and presentation of consequential data, or data that show the impact of their judgments at the item level and overall student performance.

For a given round of judgments, panelists were instructed to start at Item #1 and rate each item for Level 3:

- (MC items) What percentage of students who are *just barely* at Level 3 should answer this item correctly?
- (CR items) How many points should students *just barely* at Level 3 obtain on this item?

After all items had been rated for Level 3, the panelists would restart at item #1 and rate each item answering the questions above for students just barely at Level 4, and then repeat the process for students just barely at Level 5. A practice round of judgments was used for panelists to experience making item level judgments on a small set of items.

Prior to each round of judgments, including the practice round, the panelists were asked to complete a readiness questionnaire. This questionnaire was used to gauge panelists' readiness to complete each round, having understood the instructions provided and any feedback data provided after previous rounds of judgments (i.e., data provided after Rounds 1 and 2). If panelists had any questions or concerns leading into a judgment round, then the facilitator addressed those before the panelists could engage in the judgment process. This questionnaire was maintained as post-workshop documentation of panelists' readiness for each round during the workshop. Appendix B contains the round readiness survey form used during the workshop.

Standard Setting Rounds

There were three judgment rounds for standard setting. After each round, the panelists were provided with various forms of feedback data and given the opportunity to discuss these data with each other. After Round 1, panelists received lists of items that had the widest range of judgments for each level within the tables. In other words, a table of panelists was given lists of MC items and of CR items that had the largest range of panelists' judgments within that table, for each performance level. The lists were limited to five items for each item type for each performance level. Table 9 is a mock table of items with the greatest variability in judgments across performance levels.

Panelists used this data to discuss differences in their judgments. The discussions throughout the workshop were meant to share perspectives in expectations, not to reach consensus on judgments. When completing this task, panelists may especially take note of items that appeared across performance levels (e.g., Questions 2, 13, and 17 from Table 9).

Table 9. Mock Table of Items with Greatest Variability in Judgments

Level 3 Most Variable MC Items				Level 3 Most Variable CR Items			
Question Number	Level 3 Range	Level 3 Minimum	Level 3 Maximum	Question Number	Level 3 Range	Level 3 Minimum	Level 3 Maximum
Q17	55	25	80	Q27	2	0	2
Q13	50	25	75	Q28	2	0	2
Q2	45	25	70	Q33	2	1	3
Q18	45	25	70	Q34	2	1	3
Q19	45	25	70	Q32	1	0	1
Level 4 Most Variable MC Items				Level 4 Most Variable CR Items			
Question Number	Level 4 Range	Level 4 Minimum	Level 4 Maximum	Question Number	Level 4 Range	Level 4 Minimum	Level 4 Maximum
Q2	45	35	80	Q27	2	0	2
Q13	45	35	80	Q28	2	0	2
Q17	45	35	80	Q32	2	0	2
Q24	45	30	75	Q33	2	1	3
Q10	40	35	75	Q34	2	1	3
Level 5 Most Variable MC Items				Level 5 Most Variable CR Items			
Question Number	Level 5 Range	Level 5 Minimum	Level 5 Maximum	Question Number	Level 5 Range	Level 5 Minimum	Level 5 Maximum
Q9	45	45	90	Q37	3	2	5
Q11	45	40	85	Q28	2	0	2
Q12	45	45	90	Q29	2	0	2
Q17	45	50	95	Q30	2	0	2
Q2	40	50	90	Q33	2	2	4

Additionally, panelists were provided with empirical item difficulty data. These data reflected how well students actually did on each item from the operational test administration. Percent correct or mean score values were calculated from a representative sample of students who participated in the June 2016 test administration. These data were provided during the process to inform the panelists of their own content expectations. These data, though, reflect all levels of achievement. In other words, the empirical item difficulty data were not just on the threshold, or *just barely*, students, but on all students sampled.

Round 2 of standard setting was performed just as Round 1 had been. The difference between the two rounds was that panelists were given feedback data—item judgment variability and empirical item difficulty—and engaged in discussions over these data. Panelists were instructed to revisit their judgments from Round 1 and make a new set of judgments, keeping their judgments from Round 1 or making revisions as they felt necessary. After Round 2 judgments, panelists were provided with another set of item judgment variability data (e.g., Table 8) to discuss. The panelists discussed these data within their tables, but the facilitator led a larger group discussion on items with largest variability across the entire panel.

Also after Round 2, the facilitator provided cut scores generated from the panelists' item level judgments. Each panelist saw his/her own cut score for each performance level as well as a summary of cut scores for those at the same table. Here, panelists could compare their own cut scores to those from the overall table and consider if their cut scores matched their level of expectations. Finally, the facilitator displayed impact data, or the distribution of students among performance levels based on the panel's overall cut scores. Presenting these data during the standard setting process gave the panelists the opportunity to see the consequences of their judgments and whether these consequences fit their expectations. The panelists were reminded, however, that the data should not drive their judgments; rather, their judgments should be driven by content expectations.

Following the discussion of the Round 2 feedback data, the panelists provided one final round of judgments. This round was performed just as the previous two rounds. Once the results for Round 3 were complete, panelists were shown the final recommended cut scores and corresponding impact data. As a final task, the panelists completed a workshop evaluation that asked questions ranging from how comfortable they were with specific workshop activities to how comfortable they were with the final recommended cut scores. The workshop evaluation survey is provided in Appendix C.

Level 2 Task Force

Cut scores for Level 2 were recommended by a representative group of five panelists from the larger standard setting panel. This group met at the end of the first day of standard setting to engage in an abbreviated process of recommending cut scores. The panelists discussed threshold descriptions for Level 2 and then provided one round of judgments, using the same process as was done for the other levels earlier in the day. The recommended Level 2 cut score was provided to NYSED for further deliberation.

Cut Scores and Impact Data

Cut scores were generated after each round of judgments. The median value of the individual panelists' cut scores, per performance level, was used as the recommended cut score at the table level as well as for the overall standard setting panel. Table 10 provides a summary of the cut scores by table and overall panel for all three rounds. The most dramatic change in cut scores across all rounds was seen for Level 5. The overall cut scores for Rounds 2 and 3 were much lower than for Round 1. The cut scores for Level 3 and Level 4 did decrease from Round 1 through Round 3.

Table 10. Recommended Cut Scores Across Rounds

Table	Round 1			Round 2			Round 3		
	Level 3	Level 4	Level 5	Level 3	Level 4	Level 5	Level 3	Level 4	Level 5
1	32	55	74	29	54	72	28	54	72
2	28	48	66	26	46	64	25	47	64
3	24	41	63	25	42	61	25	43	61
4	24	45	68	20	37	58	22	40	57
Panel	27	48	69	26	46	62	25	45	64

The impact data for all three rounds of standard setting are presented in Figures 1 through 3.

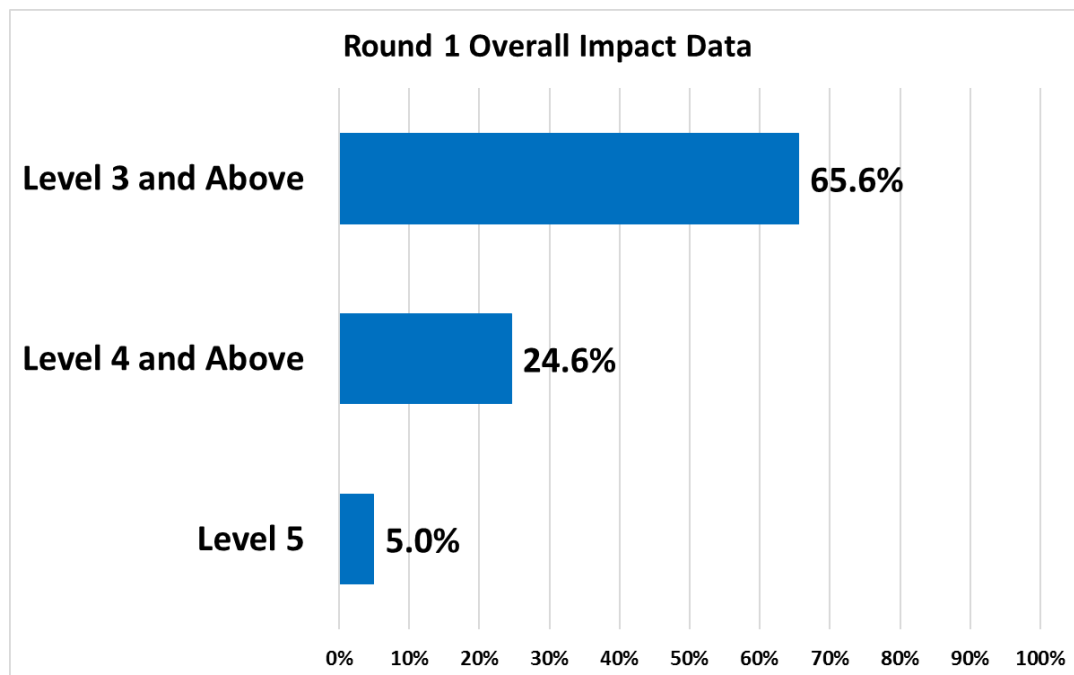


Figure 1. Round 1 Impact Data

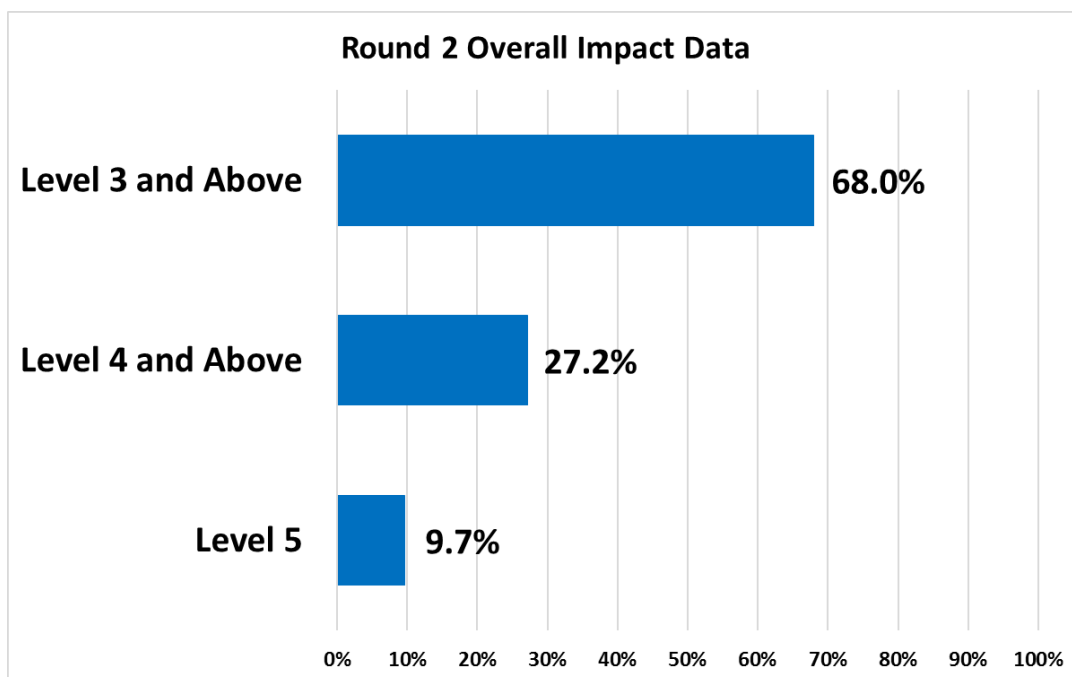


Figure 2. Round 2 Impact Data

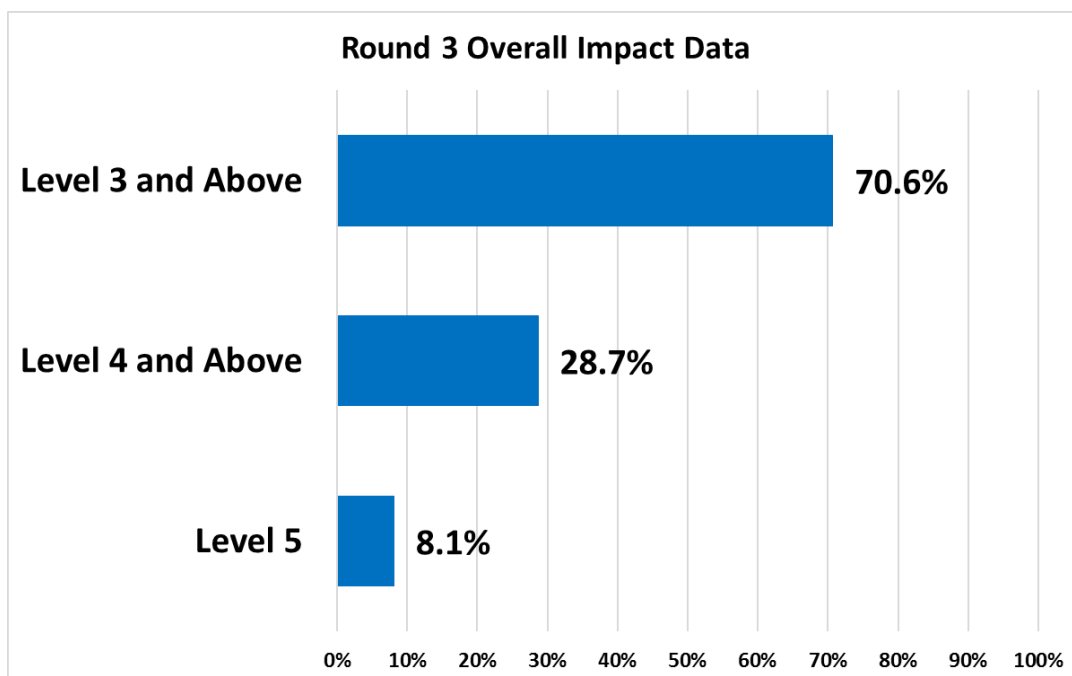


Figure 3. Round 3 Impact Data

Workshop Evaluation

Once the standard setting process was complete and the final recommended cut scores and impact data were shown, panelists completed a workshop evaluation on the various materials and activities of the standard setting process and the final recommended cut scores. The intent of this survey was to gather how well panelists understood the process, the materials used, and how comfortable they felt about the final recommended cut scores. For the survey items on the recommended cut scores,

panelists were able to express how they would modify a cut score if they were somewhat uncomfortable with the overall final recommendation. Most items on this survey used a Likert scale, with different scales of affect (e.g., strongly disagree to strongly agree, very unclear to very clear, etc.) across the evaluation.

One particular evaluation item assessed panelists' comfort in setting their individual expectations throughout the judgment process: *I rendered my judgments independently and did not feel pressured to recommend cuts at a certain score.* All of the panelists agreed to this statement, endorsing *strongly agree* or *agree*. This highlights a crucial aspect of stakeholder involvement in this high stakes activity. The 100% agreement to this particular item exhibits the support from the educators in the process of setting the performance standards for the Regents Examination in Algebra II (Common Core). In addition, some panelists provided handwritten quotes expressing their gratitude for being a part of this process and desire to participate in other educator panels for the New York State assessments.

Final Recommendations

As described in the previous sections, the NYSED, with facilitation by Pearson, conducted a formal standard setting for the Regents Examination in Algebra II (Common Core) that consisted of two meetings. The first meeting was devoted to the development of PLDs that articulated the range of knowledge and skills of students at the five levels of performance specified by New York State policy. The second meeting was dedicated to the identification of cut scores consistent with the PLDs and state policy directives, using a standardized procedure called the Modified Angoff standard setting procedure.

Both meetings reflected best practice as articulated in the Standards for Educational and Psychological Measurement (2014) and proceeded according to plans reviewed by the New York State Technical Advisory Committee. The panelists 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 examination, the rigor of the new curricula, the transitional and aspirational aspects of the New York State policy directives, and the role of the assessment in student learning throughout high school and beyond, the standard setting committee 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 Regent 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

- American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. Joint Technical Committee. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp.508–600). Washington, DC: American Council on Education.
- Egan, K. L., Schneider, M. C., and 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, NY: Routledge.
- 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., and 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, New York, NY, and available at <http://www.nciea.org>.

Appendix A: Standard Setting Workshop Agenda



Regents Examination in Algebra II (Common Core) Standard Setting

Agenda

June 15, 2016

7:30–8:00 a.m.	Registration and Breakfast
8:00–9:00 a.m.	Welcome and Standard Setting Overview
9:00–9:15 a.m.	Break
9:15–9:30 a.m.	Workshop Orientation
9:30–10:00 a.m.	Test Review
10:00–12:15 p.m.	Performance Level Descriptions
12:15–1:00 p.m.	Lunch
1:00–2:30 p.m.	Standard Setting Training
2:30–2:45 p.m.	Break
2:45–5:00 p.m.	Standard Setting <i>Round 1</i>

June 16, 2016

7:30–8:00 a.m.	Breakfast
8:00–9:45 a.m.	Standard Setting <i>Round 1</i> Feedback and Discussion
9:45–10:00 a.m.	Break
10:00–11:30 a.m.	Standard Setting <i>Round 2</i>
11:30–12:30 p.m.	Lunch
12:30–2:00 p.m.	Standard Setting <i>Round 2</i> Feedback and Discussion
2:00–2:15 p.m.	Break
2:15–3:15 p.m.	Standard Setting <i>Round 3</i>
3:15–3:45 p.m.	Break/Materials collection
3:45–4:15 p.m.	Standard Setting <i>Round 3</i> Feedback and Workshop Wrap-up

Appendix B: Panelist Readiness



Regents Examination in Algebra II (Common Core) 2016 Standard Setting

Panelist Readiness Survey Form

Panelist ID: _____

For each round, respond to the statements provided.

Practice Round		
I understand my task for the practice round.	No	Yes
I am ready to begin the practice round.	No	Yes
Round 1		
I understand my task for Round One.	No	Yes
I am ready to begin Round One.	No	Yes
Round 2		
I understand my task for Round Two.	No	Yes
I understand the feedback data that were presented from Round One.	No	Yes
I am ready to begin Round Two.	No	Yes
Round 3		
I understand my task for Round Three.	No	Yes
I understand the feedback data that were presented from Round Two.	No	Yes
I am ready to begin Round Three.	No	Yes

Appendix C: Workshop Evaluation



Regents Examination in Algebra II (Common Core) June 2016 Standard Setting

Evaluation

The purpose of this evaluation is to gather your feedback about the standard setting process. Your feedback will provide a basis for evaluating the training, methods, and materials in the standard setting process.

Please complete the information below.

- Please read each of the following statements carefully. Place an X in one box for each statement to indicate your level of agreement.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a. I understood the purpose of this standard setting workshop.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The training materials contained the information I needed to complete my assignments.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. The training on the standard setting process gave me the information I needed to complete my ratings.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. The training on the performance level descriptions gave me the information I needed to complete my assignments.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. The feedback on cut scores gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. The feedback on impact data gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. I could clearly distinguish between the performance levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h. The descriptions of performance levels were clear to me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i. I felt comfortable expressing my opinions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j. I rendered my judgments independently and did not feel pressured to recommend cuts at a certain score.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
k. I believe that the recommended cut scores represent the expectations of performance for the students that I teach.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



2. Please rate the clarity of the following materials used in the standard setting process.

	Very Unclear	Somewhat Unclear	Somewhat Clear	Very Clear
a. Instructions provided in the training materials	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Instructions provided by the facilitator	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Please rate the usefulness of the following materials or procedures in completing the standard setting process.

	Not at all useful	Somewhat useful	Very useful
a. Reviewing the assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Practicing the standard setting process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Training materials	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Table discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Large group discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. How important was each of the following in rendering your judgments?

	Not important	Somewhat important	Very important
a. Performance level descriptions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Your perception of the difficulty of the items	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Your experiences with students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Table discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Large group discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Rating feedback data (e.g., table medians)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. Impact data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



5. Were any materials or procedures especially influential in your judgments? If so, which ones? In what ways were they especially influential?

6. How appropriate was the amount of time you were given to complete the different components of the standard setting process?

	Too much	Too little	About right
a. Reviewing the assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Threshold descriptions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Training on the standard setting process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Table discussions on feedback	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Group discussions on feedback	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Please rate the extent to which you agree or disagree that the final recommended cut scores are aligned with the performance level descriptions (PLDs). *Note: Only consider the recommended cut scores for Level 3, Level 4, and Level 5.*

	Strongly Disagree	Disagree	Agree	Strongly Agree
The final recommended cut scores are aligned to the PLDs.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Please rate how comfortable you are with the final recommended cut scores. *Note: Only consider the recommended cut scores for Level 3, Level 4, and Level 5.*

	Very Uncomfortable	Somewhat Uncomfortable	Somewhat Comfortable	Very Comfortable
How comfortable are you with the final recommended cut scores?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



9. If you are not comfortable with the Level 3 cut score, would you move it (indicate with a check):

N/A, I am comfortable with the recommended Level 3 cut score: _____

Higher: _____

Lower: _____

10. If you are not comfortable with the Level 4 cut score, would you move it (indicate with a check):

N/A, I am comfortable with the recommended Level 4 cut score: _____

Higher: _____

Lower: _____

11. If you are not comfortable with the Level 5 cut score, would you move it (indicate with a check):

N/A, I am comfortable with the recommended Level 5 cut score: _____

Higher: _____

Lower: _____

12. What additional thoughts do you have on your experience participating in this standard setting workshop? (Please use the reverse side as necessary.)
