

Standard Setting Report: Illinois Assessments

Version 1.2 – November 20, 2025



Change Log

Version	Date	Name	Change
1.0	8/28/25	T. Gardner, J. Moore	Create first draft of Standard Setting Report
1.1	9/15/25	T. Gardner	Added a table with theta cuts , slope, and intercept that correspond to the new cuts scores on page 131.
1.2	10/30/25- 11/20/2025	T. Gardner, J. Moore	Addressed ISBE feedback.

Table of Contents

TABLE OF CONTENTS	3
EXECUTIVE SUMMARY	9
INTRODUCTION.....	13
Purpose and Scope	13
System Changes in 2025	13
ASSESSMENT BACKGROUND	15
Illinois Assessment of Readiness (IAR): Grades 3–8 ELA and Mathematics	15
Administration.....	15
Assessment Overviews	16
Illinois Science Assessment (ISA) – Grades 5 and 8 Science	26
Administration and Scoring.....	26
ISA Design Details.....	27
ACT Assessment: High-School ELA, Mathematics, and Science	28
Test Content.....	28
Scoring.....	29
ACT College Readiness Benchmarks	30
Summary of Assessments	36
PERFORMANCE LEVEL DESCRIPTORS	37
Development of New PLDs	37
Structure of PLDs	37
Summary PLD.....	38
Detailed PLD.....	38
STANDARD SETTING METHODOLOGY	39
Extended Modified Yes/No Angoff Method	39
Hybrid Angoff/Modified Briefing Book Method (ACT – Round 3 & 4)	41
PARTICIPANTS	48
Panelist Recruitment and Composition	48
Recruitment Process.....	48
Committee Assignments	48
Geographic and Institutional Representation	48
Panelist Demographics	49

Professional Experience	50
Panel Composition by Committee	51
Facilitators and Support Staff	52
ACT Committees.....	54
Facilitators	54
Standard Setting Roles.	55
PREPARATION.....	57
Planning and Coordination	57
Standard Setting Website and Materials	58
Materials Development Process.....	62
Facilitator Guides and Slides	63
Practice Items and Training Materials	63
Data Preparation.....	63
Vertical Articulation Tool	64
Meeting Logistics	64
Security measures	64
Contingency planning.....	65
IAR & ISA Process	65
General Session.....	66
Breakout Sessions – Training and Round 1	67
Experience the Assessment.....	67
Performance Level Descriptor (PLD) Review.....	68
Borderline Student Definition & Discussion.....	68
Methodology Training	69
Practice Judgment Activity.....	71
Readiness Quiz.....	72
Judgment Rounds for IAR/ISA Committees.....	73
Evaluations	77
ACT Standard Setting Process (Concurrent Rounds).....	78
Day 1 (Tuesday)	78
Day 2 (Wednesday)	79
Round 1 (ACT Modified Angoff Yes/No)	79
Round 2 (ACT Modified Angoff Yes/No)	82
Day 3 (Thursday)	84
Round 3 (ACT Briefing Book).....	84

Round 4 (ACT Briefing Book).....	84
Vertical Articulation (Day 5)	85
RESULTS	87
Results for Grades 3-8	87
IAR ELA	87
IAR Math	94
ISA Science	101
ACT Results for High School Grades	105
VERTICAL ARTICULATION RESULTS	112
Within-Content Vertical Articulation	112
Vertical Articulation Adjustments	113
ELA Impact Data	113
Mathematics Impact Data	115
Science Impact Data.....	117
Cross-Content Coherence.....	119
Evaluation Survey and Panelist Feedback	120
APPROVAL OF STANDARD SETTING RESULTS	123
REFERENCES	126
APPENDICES	128
Appendix A: Agendas	128
Appendix B: Score Profile Example	148
Appendix C: ACT Data Books	151
Appendix D: Evaluation 1	155
Appendix E: Evaluation 2	158
Appendix F: ACT Evaluations	160
Appendix G: Vertical Articulation Survey	165
Appendix H: Presentations	170

List of Tables

Table 1. ELA Approved Scaled Scores for Illinois Assessment for Learning (Scale = 650 to 850)	11
Table 2. Math Approved Scaled Scores for Illinois Assessment for Learning (Scale = 650 to 850)	12
Table 3. Approved Scaled Scores for Illinois Science Assessment (Scale = 700 to 900).....	12
Table 4. Approved Scaled Score Cuts for ACT.....	12
Table 5. Summary of Illinois Assessments.....	16
Table 6. Grade 3 ELA Blueprint by Claim for the Form A Standard Setting Form.....	20
Table 7. Grades 4–8 ELA Blueprint by Claim for the Form A Standard Setting Form.....	20
Table 8. PCR Task Type Max Points per Trait	22
Table 9. Calculator Use by Unit.....	23
Table 10. Summary of Mathematical Practices.....	24
Table 11. Summary of Assessment Intentions by Task Types and Subclaims.....	25
Table 12. Science Operational Test Blueprint for Grades 5 & 8.....	28
Table 13. Content Blueprint for PreACT 9 Secure, PreACT Secure, and ACT English Assessments.	32
Table 14. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Math Assessments	33
Table 15. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Reading Assessments	34
Table 16. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Science Assessments	35
Table 17. Summary of Illinois Assessments and Standard Setting Scope.....	36
Table 18. Probability of Success in College Algebra and Percentage At or Above ACT Math Score	44
Table 19. Geographic and Institutional Representation.....	49
Table 20. Panelist Demographics.....	49
Table 21. Professional Experience (N = 145 reporting).....	50
Table 22. Specialized Roles and Experience.....	50
Table 23. Committee Composition.....	51
Table 24. Design of Committees.....	52
Table 25. Lead Roles and Responsibilities for Pearson and ACT Staff.....	53
Table 26. Panelist Materials – IAR & ISA.....	59
Table 27. Meeting Agenda Overview: Modified Angoff Yes/No Approach.....	65
Table 28. Judgment Question Breakdown	70
Table 29. Questions for Extended Modified-Angoff (Yes/No) Question by Item Type	71
Table 30. Round 1 Cut Score Recommendations: Science	74
Table 31. Round 1 Sample Panelist Judgments	75

Table 32. Round 1 Sample Profile of Conditional P-Values	76
Table 33. Meeting Agenda Overview: Briefing Book Method - ACT.....	78
Table 34. Excerpt of Printout of Individual Judgments for ACT Round 1.....	79
Table 35. Round 1 Individual Raw Scores: ACT English.....	80
Table 36. Score Statistics: Scale Scores.....	82
Table 37. Excerpt of Round 1 Percent Agreement of Item Judgments: English.....	82
Table 38. Round 1 Cut Scores with Probabilities of Success and Impact: English	83
Table 39. Scaled Score Results by Round for ELA.....	89
Table 40. Scaled Score Results by Round for Math	96
Table 41. Scaled Score Results by Round for Science	102
Table 42. Round 1 Cut Score Statistics.....	105
Table 43. Round 2 Cut Score Statistics	106
Table 44. Round 3 Cut Score Statistics.....	106
Table 45. Round 4 Cut Score Statistics.....	107
Table 46. PreACT Secure Scores Linked to the Round 4 ACT Cut Scores.....	111
Table 47. Vertical Articulation Adjustments to Cut Scores: ELA	113
Table 48. Vertical Articulation Adjustments to Cut Scores: Mathematics.....	115
Table 49. Vertical Articulation Adjustments to Cut Scores: Science.....	117
Table 50. Adopted Cutscores and Cutscore Ranges for IAR and ISA.....	124
Table 51. Adopted Cutscores and Cutscore Ranges for ACT	124
Table 52. Scale Transformation Constants For Theta to Scaled Score Transformations.....	125

List of Figures

Figure 1. ELA Unit Times and High-Level Test Blueprints for Grades 3-8.	19
Figure 2. Performance Levels and Cut Scores	40
Figure 3. Plot of Probabilities of Success in College Algebra by ACT Math Score.....	45
Figure 4. Readiness Quiz	72
Figure 5. Round 3 Impact Data (Example): ELA.....	76
Figure 6. Excerpt of Conditional P-Values for ACT	80
Figure 7. Round 1 Dot Plot for ACT English.....	81
Figure 8. Round 1 Impact Data, English.....	83
Figure 9. ELA Grade 3 Panelist Agreement	90
Figure 10. ELA Grade 4 Panelist Agreement	90
Figure 11. ELA Grade 5 Panelist Agreement.....	91
Figure 12. ELA Grade Panelist Agreement.....	91
Figure 13. ELA Grade 7 Panelist Agreement.....	92
Figure 14. ELA Grade 8 Panelist Agreement.....	92
Figure 15. ELA Impact Data: Round 2	93
Figure 16. ELA Impact Data: Round 3	93
Figure 17. Mathematics Grade 3 Panelist Agreement	96
Figure 18. Mathematics Grade 4 Panelist Agreement.....	97
Figure 19. Mathematics Grade 5 Panelist Agreement.....	97
Figure 20. Mathematics Grade 6 Panelist Agreement.....	98
Figure 21. Mathematics Grade 7 Panelist Agreement	98
Figure 22. Mathematics Grade 8 Panelist Agreement.....	99
Figure 23. Math Impact Data: Round 2	99
Figure 24. Math Impact Data: Round 3	100
Figure 25. Science Grade 8 Panelist Agreement	102
Figure 26. Science Grade 8 Panelist Agreement	103
Figure 27. Science Impact Data: Round 2.....	103
Figure 28. Science Impact Data Round 3.....	104
Figure 29. Panelist Agreement: Math	107
Figure 30. Panelist Agreement: Science	108
Figure 31. Panelist Agreement: ELA	108
Figure 32. Round 4 Impact Data: Percentage of Illinois Students Scoring Within Each Performance Level.....	109
Figure 33. ELA Impact Data by Performance Level After Vertical Articulation.....	114
Figure 34. ELA Impact Data by Proficient vs. Non-Proficient After Vertical Articulation	114
Figure 35. Mathematics Impact Data by Performance Level After Vertical Articulation.....	116
Figure 36. Mathematics Impact Data by Proficient vs. Non-Proficient After Vertical Articulation.....	116
Figure 37. Science Impact Data by Performance Level After Vertical Articulation.....	118
Figure 38. Science Impact Data by Proficient vs. Non-Proficient After Vertical Articulation	118
Figure 39. Percentage of Proficient Students in each Grade by Subject.....	120

Executive Summary

On July 14–18, 2025, ISBE convened 19 educator committees in Springfield to recommend cut scores that define four performance levels for the Illinois Assessment of Readiness - IAR (Grades 3–8 ELA & Mathematics), the Illinois Science Assessment - ISA (Grades 5 & 8 Science), and the ACT (Grade 11 ELA, Mathematics, and Science). Panels used the Extended Modified Yes/No Angoff method for IAR/ISA. For ACT, panels completed Angoff rounds followed by data-informed “Briefing Book” rounds to align the Grade 11 standards with college-readiness evidence. Vertical articulation on Day 5 reviewed progression across grades and ensured coherence within and across content areas. After Board approval on August 13, 2025, the new cuts were adopted for 2025 reporting and define the four performance levels statewide.

Committees and Methods. Nineteen committees (14 IAR/ISA; 5 ACT subject tests) conducted multiple Angoff rounds with feedback and discussion; the ACT committees additionally completed Briefing Book rounds using Illinois 2025 ACT data, PreACT Secure data (Grades 9–10), and multi-state census data to contextualize cuts against external indicators (benchmarks, probabilities of A/B/C in first-year courses). Final committee medians served as recommended cuts. ACT Grade 11 cuts were statistically linked to PreACT Secure and PreACT 9 Secure using equipercentile methods to maintain coherent high-school expectations.

Panel Composition. ISBE selected 155 panelists from over 900 applicants, representing all seven regions and 45+ counties; detailed demographics and experience are reported in the Participants section. The panelists were predominantly current or recently practicing educators with deep content knowledge and experience instructing diverse student populations. On average, panelists had 17 years of teaching experience (with a range of 2 to 42 years) and had demonstrated familiarity with the Illinois Learning Standards and the state assessments. Efforts were made to ensure a balanced representation of educators across grade levels, school types, and demographic backgrounds, reflecting Illinois’s K–12 population. All panelists and staff convened in person under secure conditions, supported by trained facilitators, psychometricians, content experts, program management, and observers from ISBE, Pearson, and ACT.

IAR & ISA Standard Setting Rounds. Over the five-day meeting, panelists engaged in a structured, iterative judgment process to recommend cut scores for each performance level. They were first trained on the [performance level descriptors](#) (PLDs) defining the knowledge and skills of students at each level, and they practiced applying a borderline student concept to the test content. Using the Extended Modified Yes/No Angoff method, the grade-level committees conducted three rounds of judgments for the IAR and ISA assessments. After each round, facilitators provided feedback data, including summaries of panelists’ initial cut score recommendations, group statistics, and, in later rounds,

projected impact data (the percentage of students statewide who would fall into each performance level given the current cut scores), to inform panelists' discussions and subsequent ratings. Panelists discussed any large discrepancies in their item ratings or cut score selections and were encouraged to consider their colleagues' rationales, though not required to reach consensus. Through this iterative process, the variability in cut score judgments typically decreased and the panel moved toward convergence by Round 3, indicating growing agreement on the performance standards.

ACT Standard Setting Rounds. The ACT standard setting followed a hybrid two-phase process. In Rounds 1 and 2, five subject-specific committees (ACT English, Mathematics, Reading, Science, and Writing) applied the Modified Yes/No Angoff approach, reviewing each test item and indicating whether a borderline student at each cut score level would likely answer it correctly (for the writing test, panelists indicated the score point representing the most common response for students at the lower end of each proficiency level). This provided content-based initial cut score recommendations for each of the five ACT subject tests. Before Rounds 3 and 4, the ACT panelists were provided briefing books containing a variety of empirical data, including performance of multiple ACT-tested populations, percentages of students meeting the ACT College Readiness Benchmarks, probabilities of success in college courses by ACT score, college enrollment rates by ACT score, Illinois performance on the PreACT Secure assessments, and historical performance data. The ACT panelists then reconvened in three groups corresponding to the ACT ELA, Math, and Science score reporting (with English, Reading, and Writing combined into an overall ELA score). In these final two rounds, the panelists used the empirical data to adjust their cut score recommendations, focusing on aligning performance standards with indicators of college and career readiness. Discussions took place both in subject-specific breakouts and in combined sessions to balance content considerations with the external data. By the end of Round 4, the ACT committees had recommended cut scores for the ACT ELA, ACT Mathematics, and ACT Science scales. These ACT cut scores were then used as the basis for determining equivalent cut scores on the PreACT 9 Secure (grade 9) and PreACT Secure (grade 10) assessments via statistical linking, ensuring a vertically aligned progression of performance standards from Grade 9 through 11.

Vertical Articulation & Coherence. On the final day, vertical articulation sessions were held for ELA, Mathematics, and Science. Each vertical articulation committee included 2–4 panelists from each grade-level committee in that content across grades 3–11. The ELA and Mathematics committees had 14 panelists, representing grades 3–11, and the Science committee had 6 panelists, representing grades 5, 8, and 11. In these sessions, panelists reviewed the recommended cut scores across all grades as a complete continuum. Using an impact data calculator tool and cross-grade performance data, they examined whether the stringency of “Proficient,” for example, increased appropriately from one grade to the next. Committees reviewed grade-to-grade progressions to resolve anomalies (e.g., inversions or disproportionate jumps) so that PLDs and cuts reflect steadily increasing

expectations from Grades 3–8 (IAR/ISA) to Grade 11 (ACT). All panelists from the Vertical Articulation committees came together for a final content coherence discussion where panelists voted on the final cuts.

This Vertical Articulation section of this report provides the final recommended cuts by grade/subject and the corresponding impact distributions by performance level and by proficient/non-proficient. It also summarizes panelist agreement statistics across rounds, participant demographics, and evaluation survey results. (All panelists completed evaluation surveys at the end of their committee work, providing feedback on the clarity, fairness, and effectiveness of the standard setting process. The survey responses indicated high overall satisfaction with the process and confidence in the recommended cut scores - detailed in a later section of this report). ISBE staff and observers noted that panelists were deeply engaged and thoughtful throughout the process, and that the facilitated discussions were collaborative and professional. The outcome of this effort is a set of recommended performance standards that are vertically aligned (Grades 3–11) and anchored in college and career readiness, and were presented and approved by the Illinois State Board of Education with no adjustments. After Board approval the cut scores were adopted for reporting student results on the 2025 IAR, ISA, PreACT Secure, and ACT assessments, defining the four new performance levels for Illinois students. The final scaled score cuts that were adopted by the Board in August, 2025 are displayed in Tables 1–4 below. For reference the former cuts are also presented along with the difference between the former cuts and current cuts.

Table 1. ELA Approved Scaled Scores for Illinois Assessment for Learning (Scale = 650 to 850)

Grade Level	Approaching Proficient		Proficient		Above Proficient	
	Former	Current	Former	Current	Former	Current
3	700	685 (-15)	750	735 (-15)	810	780 (-30)
4	700	695 (-5)	750	737 (-13)	790	780 (-10)
5	700	700 (0)	750	739 (-11)	799	780 (-19)
6	700	705 (+5)	750	741 (-9)	790	780 (-10)
7	700	710 (+10)	750	743 (-7)	785	785 (0)
8	700	710 (+10)	750	745 (-5)	794	795 (+1)

Table 2. Math Approved Scaled Scores for Illinois Assessment for Learning (Scale = 650 to 850)

Grade Level	Approaching Proficient		Proficient		Above Proficient	
	Former	Current	Former	Current	Former	Current
3	700	705 (+5)	750	732 (-18)	790	781 (-9)
4	700	708 (+8)	750	740 (-10)	790	784 (-6)
5	700	709 (+9)	750	740 (-10)	790	782 (-8)
6	700	705 (+5)	750	742 (-8)	785	773 (-12)
7	700	712 (+12)	750	745 (-5)	785	781 (-4)
8	700	705 (+5)	750	745 (-5)	785	791 (+6)

Table 3. Approved Scaled Scores for Illinois Science Assessment (Scale = 700 to 900)

Grade Level	Approaching Proficient		Proficient		Above Proficient	
	Current	Recommended	Current	Recommended	Current	Recommended
5	770	770 (0)	799	812 (+13)	832	856 (+23)
8	770	770 (0)	799	812 (+13)	832	856 (+23)

Table 4. Approved Scaled Score Cuts for ACT

Grade Level	Approaching Proficient			Proficient			Above Proficient		
	ELA	Math	Science	ELA	Math	Science	ELA	Math	Science
9	11	15	11	14	17	14	25	24	25
10	12	15	11	15	18	16	26	25	25
11	13	15	14	18	19	19	27	28	27

Grade 9 (PreACT 9 Secure) Scale 1-32
 Grade 10 (PreACT Secure) Scale 1-35
 Grade 11 (ACT with Writing) Scale 1-36

Introduction

Purpose and Scope

This technical report documents Illinois’s July 2025 standard setting for IAR (Grades 3–8 ELA & Math), ISA (Grades 5 & 8 Science), and ACT (Grade 11 ELA, Math, Science). Educator panels reviewed PLDs and test content, applied the Extended Modified Yes/No Angoff method (IAR/ISA), and— for ACT—combined Angoff with Briefing Book rounds anchored in college-readiness evidence to recommend cut scores that define four performance levels.

System Changes in 2025

Illinois transitioned high-school accountability from SAT to ACT and aligned all assessments to a four-level structure: Level 1 (Below Proficient), Level 2 (Approaching Proficient), Level 3 (Proficient), Level 4 (Above Proficient). The four-level change unified reporting across IAR, ISA, and ACT and improved coherence from elementary through high school.

The purpose of this technical report is to provide a comprehensive account of how the new performance standards were established. The report is organized into several sections:

1. The **Assessment Background** section provides an overview of each assessment (IAR, ISA, and ACT) including their content, design, and scoring, to contextualize the standard setting.
2. The **Performance Level Descriptors** section describes the development and role of the PLDs, which articulate what students at each performance level know and can do.
3. The **Standard Setting Methodology** section outlines the specific methods used (Extended Modified Angoff and the hybrid Briefing Book approach for ACT) and the rationale behind them.
4. The **Participants** section describes the composition and recruitment of the standard setting panel and other personnel (facilitators, analysts, etc.) involved.
5. The **Preparation** section summarizes the steps taken prior to the meeting to prepare materials, data, and the meeting logistics.
6. The **Process** section details the execution of the standard setting meeting itself, including training activities, breakout sessions, judgment rounds, feedback and discussions, and the vertical articulation process. Within the Process description,

distinctions are made between the procedure followed for IAR/ISA (which relied solely on the Modified Angoff rounds) and the procedure for ACT (which incorporated empirical data in later rounds). The Evaluation Survey Summary provides an overview of panelists' feedback on the process.

7. The **Results** section presents the outcomes of the meeting – namely the recommended cut scores for each grade and subject – along with placeholders for tables showing detailed results such as round-by-round cut score data, impact data, and other analyses.
8. Finally, the **Vertical Articulation** section discusses how the cut scores were reviewed holistically across grades and subjects and any adjustments that resulted from that review. References to relevant literature and appendices (e.g., agendas, detailed data tables, panelist rosters) are included at the end of the report.

In summary, Illinois's 2025 standard setting was a collaborative, research-based endeavor to establish defensible performance standards that are aligned across a K–12 continuum and tied to indicators of college and career readiness. The following sections document each aspect of this endeavor, ensuring transparency in how the cut score recommendations were derived and providing evidence to support the validity of the results.

Assessment Background

Illinois Assessment of Readiness (IAR): Grades 3–8 ELA and Mathematics

The Illinois Assessment of Readiness (IAR) is the state’s annual accountability assessment in English Language Arts (ELA) and Mathematics for students in Grades 3 through 8. The IAR is aligned to the Illinois Learning Standards, which incorporate the Common Core State Standards for ELA and math. It is administered each spring to all public school students in those grades. The IAR replaced the PARCC assessment in 2019 and has since been Illinois’s measure of students’ mastery of grade-level content in ELA and Mathematics.

The IARELA and math tests are criterion-referenced and built to assess the breadth of the Illinois Learning Standards at each grade. Each test is composed of a mix of item types, including traditional multiple-choice questions and technology-enhanced items. In ELA, students typically encounter passages with associated questions, including evidence-based selected response items and prose constructed response (essay) items that assess writing skills. In math, items cover major content domains (such as algebra, geometry, number & operations, etc.) as well as mathematical practice skills like problem solving and reasoning. Calculators are permitted on certain sections of the math test in accordance with the test’s blueprint (for example, on some units for grades 6–8, while not allowed on others) – this ensures that both computational fluency and conceptual understanding are assessed.

Each IAR test yields a scaled score that is specific to grade and subject. The score scales differ by grade, but all are structured such that higher scores indicate better performance. Before 2025, the IAR reported performance in five levels. As of 2025, the IAR reports four performance levels (Level 1 through 4 as described above). One goal of the standard setting was to determine the new cut scores on each grade’s IAR ELA and math scale that separate these four levels.

Administration

The IAR is administered primarily online via Pearson’s TestNav platform, though paper forms are available as accommodations. The test is timed and taken in multiple units over a few days. The IAR is designed to provide reliable and valid scores at the student, school, and district levels, and it serves as one measure for federal accountability. Because the IAR includes a variety of item formats (including essays in ELA and multi-part tasks in math), raw scores are calculated as a combination of machine-scored and human-scored items.

These raw scores are converted to the scaled score through psychometric equating, ensuring that scores are comparable across different test forms and years.

Assessment Overviews

This section describes the high-level overviews for the Illinois Assessment of Readiness (IAR) for ELA & Mathematics for grades 3–8, the Illinois Science Assessment (ISA) for grades 5 & 8, and ACT Assessments for high school for grades 9–12.

Table 5. Summary of Illinois Assessments

Assessment	Grades										
	3	4	5	6	7	8	9	10	11		
Illinois Assessment for Readiness (IAR) – ELA & Math											
Illinois Science Assessment (ISA) – Science											
ACT Suite – English, Reading, Math, Science, & *Writing											

*Writing is only included on the ACT at grade 11. It is not included on the PreACT 9 Secure or PreACT Secure at grade 10.

IAR ELA High-Level Design

The ELA summative assessment is available in Grades 3 – 8. Students read and analyze passages from grade-level texts including media, with a range of complexities– both fiction and nonfiction – as well as multimedia. Students then draw upon what they’ve learned from these passages and multimedia, using that knowledge to respond in writing. These skills are critically important for students in college and the workplace.

The IARELA claim structure ensures that students are evaluated on their ability to read and comprehend both literary and informational texts independently, understand and use vocabulary in context, and write effectively with proper language conventions.

ELA/Literacy Master Claim

Students must demonstrate that they are on track to readiness by reading and comprehending grade-level texts of appropriate complexity and writing effectively when using and/or analyzing sources.

ELA Major Claims & Sub-Claims

1. Reading: Students should be able to read and comprehend a range of sufficiently complex texts independently.
 - a. **Reading – Vocabulary Interpretation and Use (RV):** In this subclaim, students are assessed on their ability to understand and interpret vocabulary within the context of a text. This involves identifying the meanings of words and phrases, including figurative and connotative meanings, and analyzing the impact of specific word choices on the text's meaning and tone. Students must demonstrate the ability to use context clues, word structures like prefixes and suffixes, and reference materials to determine or clarify the meaning of unfamiliar words, ensuring they can fully comprehend the text.
 - b. **Reading – Reading Literature (RL):** This subclaim focuses on students' proficiency in reading and understanding complex literary texts. Students are expected to analyze themes, settings, and character development, and to understand how narrative techniques and points of view shape the story. Additionally, they must interpret literary elements such as symbolism, metaphors, and irony, and evaluate how different components of a literary work contribute to its overall meaning and impact.
 - c. **Reading – Reading Informational Text (RI):** This subclaim assesses students' ability to read and comprehend informational texts like essays, articles, and reports. Students are required to analyze the development of central ideas or arguments, evaluate the structure and use of text features, and assess the effectiveness of the reasoning and evidence presented. Understanding the author's purpose and perspective, as well as the rhetorical techniques used to influence the reader, are key aspects of this subclaim.
2. Writing: Students should be able to write effectively when using and/or analyzing sources.
 - a. **Writing – Written Expression (WE):** Students are evaluated on their ability to express their ideas clearly and coherently in writing. This includes developing and organizing their writing according to the task, purpose, and audience, with a focus on crafting clear thesis statements and a logical progression of ideas. Students must also support their claims or narratives with relevant

evidence and details, while effectively structuring their writing with strong introductions, conclusions, and transitions.

- b. **Writing – Knowledge of Language and Conventions (WKL):** This subclaim focuses on students' command of standard English conventions in their writing. Students must demonstrate correct use of grammar, punctuation, capitalization, and spelling, as well as standard usage rules such as subject-verb agreement and consistent verb tenses. Additionally, they are expected to vary sentence structures appropriately and make syntactical choices that enhance clarity and precision, ensuring their writing is coherent and easy to read.

Additional ELA Resources: Task Types & Item Types

The Illinois Assessment of Readiness (IAR) for ELA uses a variety of documents to structure and define the assessment process. Evidence Statement Tables are used to specify the knowledge and skills students are expected to demonstrate on the assessment, aligning each item with specific standards and sub-claims. The Test Blueprints provide a detailed overview of the content and structure of the assessment, outlining the number of items, point values, and the distribution of questions across various content areas and skill sets. Performance Level Descriptors (PLDs) define the knowledge, skills, and practices that students at different performance levels are expected to demonstrate, helping to categorize their proficiency as below, approaching, meeting, or exceeding expectations.

The IARELA assessment employs several item types to assess student understanding. The Evidence-Based Selected Response (EBSR) requires students to select answers based on evidence from the text, often involving two-part questions where the first part asks for an answer and the second part asks for the evidence supporting that answer. Technology-Enhanced Constructed Response (TECR) items require students to interact with the test content in digital formats, such as dragging and dropping, or selecting multiple pieces of evidence. Prose Constructed Response (PCR) tasks require students to write extended responses that demonstrate their ability to analyze texts, synthesize information, or develop narratives.

Scoring Rubrics for these tasks measure different dimensions of student responses. For each of the Task Model Types, specific rubrics are applied. The Literary Analysis Task (LAT) requires students to read two literature passages and answer associated questions, culminating in a PCR task where they analyze the texts. The Research Simulation Task (RST) involves reading 2–3 informational pieces, answering related questions, and completing a PCR task focused on synthesizing information from the texts. The Narrative Writing Task (NWT) presents students with one literature passage, followed by questions and a PCR task where they must create a narrative. Each task type is measured using rubrics that typically

evaluate dimensions like content understanding, use of evidence, and writing conventions, with points allocated according to the complexity and depth of the response.

Lastly, the test blueprint also includes a Short Passage Set and a Medium/Long Passage Set, which consist of a range of texts accompanied by questions designed to assess students' comprehension and analytical abilities, but do not contain the PCR writing task like the task model described above. These passage sets allow for additional reading items on the assessment in the Reading for Literature and Reading for Information subclaims without the additional testing time that is required for an additional writing prompt. Links to some of these additional resources can be found below along with a bulleted summary. The highlevel test blueprint is included in the following figure.



ELA Unit Times & Blueprints

GRADE 3 FORM A

Unit 1	Literary Analysis Task	75 min
Unit 2	Research Simulation Task	75 min
		2.5 HRS Total

GRADES 4 – 8 FORM A

Unit 1	Literary Analysis Task	90 min
	Short Passage	
Unit 2	Research Simulation Task	90 min
		3 HRS Total

GRADE 3 FORM B

Unit 1	Narrative Writing Task	75 min
	Short Passage	
Unit 2	Research Simulation Task	75 min
		2.5 HRS Total

GRADES 4 – 8 FORM B

Unit 1	Narrative Writing Task	90 min
	Long or Paired Passage	
Unit 2	Research Simulation Task	90 min
		3 HRS Total

Students are assigned either Form A or Form B at random

Figure 1. ELA Unit Times and High-Level Test Blueprints for Grades 3-8.

The tables below provide more detail for the grade level blueprints. Grades 4-8 have a common test structure, so that information is combined in the same table.

Table 6. Grade 3 ELA Blueprint by Claim for the Form A Standard Setting Form

Task/Item Set	Number of Passages	Claims/Sub-Claims	Max Points from EBSR/TECR Items	Max Points from PCRs
Unit 1				
Literary Analysis Task	2	Reading: Literary Text	8	3
		Reading: Vocabulary	4	0
		Writing: Written Expression	0	9
		Writing: Knowledge of Language and Conventions	0	3
Unit 2				
Research Simulation Task	2	Reading: Informational Text	8	3
		Reading: Vocabulary	4	0
		Writing: Written Expression	0	9
		Writing: Knowledge of Language and Conventions	0	3
Totals	4		24 Reading	6 Reading 24 Writing

Table 7. Grades 4-8 ELA Blueprint by Claim for the Form A Standard Setting Form

Task/Item Set	Number of Passages	Claims/Sub-Claims	Max Points from EBSR/TECR Items	Max Points from PCRs
Unit 1				
Literary Analysis Task	2	Reading: Literary Text	8	4
		Reading: Vocabulary	4	0
		Writing: Written Expression	0	12
		Writing: Knowledge of Language and Conventions	0	3
Short Passage Set	1	Reading: Literary Text	6	N/A
		Reading: Vocabulary	2	N/A
Unit 2				
Research Simulation Task	3	Reading: Informational Text	12	4
		Reading: Vocabulary	4	0
		Writing: Written Expression	0	12
		Writing: Knowledge of Language and Conventions	0	3
Totals	6		36 Reading	8 Reading 30 Writing

ELA Key Terms and Resources in Standard Setting Meeting

1. Evidence Statements: See grade [3](#), [4](#), [5](#), [6](#), [7](#), [8](#)
2. Test Blueprints
3. [Performance Level Descriptors](#)
4. Task Models (See: [Grades 3-5](#) & [Grades 6-8](#))
5. Item Types
 - a. Evidence-Based Select Response (EBSR)
 - b. Technology-Enhanced Constructed Response (TECR)
 - c. Prose Constructed Response (PCR)
6. [Scoring Rubrics](#)
7. Task Types
 - a. Task Model
 - i. Literary Analysis Task (LAT): 2 literature passages + associated items + PCR task
 - ii. Research Simulation Task (RST): 2-3 informational pieces + associated items + PCR task
 - iii. Narrative Writing Task (NWT): 1 literature passage + associated items + PCR task
 - b. Traditional Passage Sets
 - i. Short Passage
 - ii. Medium/Long Passage

Each test form has two operational PCR items out of the three task model types: Language Analysis Task (LAT), Research Simulation Task (RST), and Narrative Writing Task (NWT). As defined at the rubric link above, each PCR has two traits. Below are the rules for deriving maximum PCR score points. Please note that the Reading Comprehension and Written Expression trait has a maximum score of 3 or 4 points and is weighted by 3 or 4 depending on the grade. Therefore, it has a maximum score of 9 to 16 depending on the grade and task type. Knowledge of Language and Conventions is always weighted by 1 and has a max score of 3, so all grades and task types have a max score of 3 points for this subclaim. The maximum points for the PCR ranges from 12 to 19.

Table 8. PCR Task Type Max Points per Trait

GD	Task	Max Pts	Claim	Subclaim/ Points	Max Score	Max Pts. per Subclaim	Notes
3	LAT	15	Reading	Reading Comprehension + Written Expression	3	12*	RI/RL: 3 WE: 9 (3x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	RST	15	Reading	Reading Comprehension + Written Expression	3	12*	RI/RL: 3 WE: 9 (3x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	NWT	12	Writing	Written Expression	3	9*	WE: 9 (3x3)
				Writing Knowledge and Conventions	3	3	WKL: 3
4-5	LAT	19	Reading	Reading Comprehension + Written Expression	4	16*	RI/RL: 4 WE: 12 (4x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	RST	19	Reading	Reading Comprehension + Written Expression	4	16*	RI/RL: 4 WE: 12 (4x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	NWT	12	Writing	Written Expression	3	9*	WE: 9 (3x3)
				Writing Knowledge and Conventions	3	3	WKL: 3
6-8	LAT	19	Reading	Reading Comprehension + Written Expression	4	16*	RI/RL: 4 WE: 12 (4x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	RST	19	Reading	Reading Comprehension + Written Expression	4	16*	RI/RL: 4 WE: 12 (4x3)
			Writing	Writing Knowledge and Conventions	3	3	WKL: 3
	NWT	15	Writing	Written Expression	4	12*	WE: 12 (4x3)
				Writing Knowledge and Conventions	3	3	WKL: 3

IAR Mathematics High-Level Design

The mathematics summative assessment is available in Grades 3 – 8. Students solve multi-step math problems that require reasoning and address real-world situations. This requires students to reason mathematically, make sense of quantities and their relationships to solve real-world problems, and show their understanding.

The Illinois Assessment of Readiness (IAR) for Mathematics evaluates students' mastery of essential mathematical content across various domains. The Mathematics Content Assessed includes five key areas: Number, Operations and Algebraic Thinking, Measurement, Geometry, and Data, Statistics, and Probability. These areas are foundational to the development of mathematical skills necessary for academic success and future readiness.

Additional Resources for additional details:

1. Test Blueprints
2. Evidence Statements: Grades [3](#), [4](#), [5](#), [6](#), [7](#), [8](#)
3. [Performance Level Descriptors](#)

The IAR Mathematics Test Structure outlines how the test is organized, with specific guidelines for Calculator Use across different grades. For students in grades 3-5, the entire assessment is taken without a calculator, ensuring that students demonstrate their arithmetic skills independently. In grade 6, students begin to use calculators in sections 2 and 3, with a combination of non-calculator and calculator items in Unit 1. By grades 7 and 8, students primarily use calculators in Units 2 and 3, with some non-calculator items still present in Unit 1, reflecting the increasing complexity of the content and the need for computational tools.

Table 9. Calculator Use by Unit

Grades	Unit 1	Unit 2	Unit 3
3-5	Non-Calculator	Non-Calculator	Non-Calculator
6	Non-Calculator and Calculator	Calculator	Calculator
7	Non-Calculator and Calculator	Calculator	Calculator
8	Non-Calculator	Calculator	Calculator

The Mathematics Practices are a set of eight practices that guide students in applying their mathematical knowledge effectively. These practices include:

1. Making sense of problems and persevering in solving them.
2. Reasoning abstractly and quantitatively.
3. Constructing viable arguments and critiquing the reasoning of others.
4. Modeling with mathematics.
5. Using appropriate tools strategically.
6. Attending to precision.
7. Looking for and making use of structure.
8. Looking for and expressing regularity in repeated reasoning.

These practices are embedded throughout the assessment, encouraging students to think critically and apply their knowledge in meaningful ways.

Table 10. Summary of Mathematical Practices

Mathematical Practice	Description
MP.1: Make sense of problems and persevere in solving them	<ul style="list-style-type: none"> Students approach problems with understanding and determination, persisting in finding solutions even when faced with challenges.
MP.2: Reason abstractly and quantitatively	<ul style="list-style-type: none"> Students make sense of quantities and their relationships in various contexts. They understand the meaning of quantities and use them to solve problems.
MP.3: Construct viable arguments and critique the reasoning of others	<ul style="list-style-type: none"> Students justify their reasoning and critique the reasoning of others. They construct logical arguments and evaluate the arguments of others.
MP.4: Model with mathematics	<ul style="list-style-type: none"> Students apply mathematical concepts to solve real-world problems, representing situations mathematically and using mathematical models to analyze phenomena.
MP.5: Use appropriate tools strategically	<ul style="list-style-type: none"> Students select and use appropriate tools, including technological tools, to solve mathematical problems efficiently and accurately.
MP.6: Attend to precision	<ul style="list-style-type: none"> Students communicate precisely using mathematical language and notation, specifying units of measure, describing reasoning accurately, and giving precise answers.
MP.7: Look for and make use of structure	<ul style="list-style-type: none"> Students recognize and use the underlying structure of mathematical concepts to solve problems, identifying patterns, properties, and relationships.
MP.8: Look for and express regularity in repeated reasoning	<ul style="list-style-type: none"> Students identify and generalize patterns in their reasoning, looking for shortcuts and efficient strategies based on repeated reasoning, leading to generalizations and deeper understanding.

Mathematics Claim Structure: The IAR Mathematics Claim Structure focuses on whether students are "On-Track" for college and career readiness. The Major Claim assesses the degree to which a student is prepared to solve grade-level problems, making connections to the Standards for Mathematical Content and the Standards for Mathematical Practice. This is broken down into four sub-claims:

1. **Sub-Claim A: Major Content with Connections to Practices** focuses on the primary content areas for the grade level, ensuring that students can solve problems that integrate major mathematical concepts with relevant practices.
2. **Sub-Claim B: Additional & Supporting Content with Connections to Practices** targets content that supports and extends the major concepts, providing a well-rounded assessment of a student's mathematical abilities.
3. **Sub-Claim C: Mathematical Reasoning** evaluates a student's ability to construct viable arguments, critique the reasoning of others, and demonstrate precision in mathematical statements. This sub-claim is critical in assessing how well students understand and can articulate mathematical reasoning.
4. **Sub-Claim D: Mathematical Modeling** assesses the ability to solve real-world problems by applying grade-appropriate mathematical knowledge and skills, with specific focus on practices such as modeling with mathematics and using tools strategically.

Mathematics Task Types are categorized into three types of items, each designed to assess different aspects of mathematical understanding:

1. **Type 1** items are focused on the application of procedural skills and fluency. These tasks typically involve straightforward calculations or the application of specific mathematical procedures. Type 1 items are worth 1–2 points each.
2. **Type 2** items assess mathematical reasoning. These tasks require students to demonstrate their ability to reason mathematically, construct arguments, and critique the reasoning of others. Type 2 items are more complex and are worth 3–4 points.
3. **Type 3** items are focused on mathematical modeling. These tasks present students with real-world problems that require them to apply their mathematical knowledge in a practical context. Type 3 items are often multi-step and are worth 3 or 6 points, reflecting their complexity and the depth of understanding required.

The design of Subclaims A, B, C, and D is closely aligned with the three types of tasks. Sub-Claim A primarily aligns with Type 1 items, focusing on the major content areas and procedural skills. Sub-Claim B also aligns with Type 1 items but targets additional and supporting content. Sub-Claim C aligns with Type 2 items, emphasizing mathematical reasoning, while Sub-Claim D aligns with Type 3 items, focusing on mathematical modeling and the application of mathematics in real-world contexts. Together, these subclaims and task types provide a comprehensive assessment of students’ mathematical abilities, ensuring they are on track for college and career readiness.

Table 11. Summary of Assessment Intentions by Task Types and Subclaims

Type	Assessment Intention	Points
Type 1 (Subclaims A-B)	Demonstrate understanding. • Example: Define a term. Apply skills and procedures. • Example: Provide or identify an answer.	1 or 2
Type 2 (Subclaim C)	Demonstrate reasoning. • Example: Show or describe solution steps. Provide justification. • Example: State why steps are valid. Critique reasoning or justification. • Example: Identify or correct error(s) in someone’s work.	3 or 4 At least 2 points come from reasoning.
Type 3 (Subclaim D)	Demonstrate modeling. • Example: Create equation(s) for solving a problem. Provide application related to a context. • Example: State why equation(s) represent solving the problem. Critique modeling or application. • Example: Explain why someone’s model will not solve the problem.	3 or 6 At least 2 or 3 points come from modeling.

Illinois Science Assessment (ISA) – Grades 5 and 8

Science

The Illinois Science Assessment (ISA) is the statewide science assessment administered to students in Grade 5 and Grade 8 (and in Grade 11, prior to 2025 – see ACT Science below). The ISA is aligned to the Illinois Learning Standards for Science, which incorporate the Next Generation Science Standards (NGSS). The NGSS emphasize application of scientific concepts and practices, and the ISA is designed to measure students’ understanding of the core ideas in physical science, life science, Earth/space science, as well as their use of science and engineering practices.

The ISA is delivered online and consists of a series of scenario-based sets of questions. It includes multiple-choice and technology-enhanced items, as well as some constructed response items requiring short explanations. The test covers three NGSS content domains (Life, Physical, Earth/Space sciences) and integrates the NGSS crosscutting concepts and practices. At Grade 5, the content reflects standards taught in grades 3–5; at Grade 8, it reflects standards from grades 6–8 (middle school science). Each test is un-timed, though most students complete it in approximately 2 hours. The operational test blueprint ensures coverage of all science domains; for example, each form might include a balanced number of items for life science, physical science, and Earth science, possibly presented in clustered item sets that all relate to a common stimulus or experimental scenario.

Like the IAR, the ISA yields a grade-specific scaled score for each student. The score scale differs for Grade 5 and Grade 8. Prior to 2025, the ISA also used a four level performance scheme, with performance level labels of Emerging, Developing, Proficient, and Exemplary. As of spring 2025, the ISA likewise transitioned to consistent performance level labels (Below, Approaching, Proficient, Above Proficient) with the IAR. The standard setting panels for science (one for Grade 5, one for Grade 8) were tasked with recommending the new cut scores on the Grade 5 ISA scale and the Grade 8 ISA scale for those performance levels.

Administration and Scoring

The ISA is administered in the spring, typically in March. It is delivered via the same online platform as IAR, and utilizes technology features for interactive items (for example, drag-and-drop classification of organisms, selecting parts of a diagram, etc.). Students’ responses are machine-scored except for any short written responses, which are scored by trained human raters using rubrics. Scaled scoring and equating methods are used to maintain year-to-year consistency. Student results are reported with a scale score and a performance level designation. Additionally, subscore information by science domain is provided to give insight into areas of strength or weakness (e.g., Life Science subscore).

The interpretation of the new performance levels for ISA will align with the descriptions in the PLDs. For instance, a student at Level 3 (Proficient) in Grade 8 Science would be interpreted as demonstrating the science knowledge and reasoning skills expected by the end of 8th grade, such as explaining phenomena or interpreting data consistent with NGSS expectations for middle school. The cut scores determined through standard setting will ensure that such interpretations are grounded in actual performance data and educator judgments about what a minimally Proficient student can do.

ISA Design Details

The ISA supports educational instruction and assesses student learning by providing feedback to educators and parents based on the Illinois Learning Standards for Science incorporating the Next Generation Science Standards (NGSS), which can be used to inform instructional strategies, to enrich curriculum, and to remediate learning. More information about the administration of the ISA can be found in the ISA Test Administrator Manual. The Next Generation Science Standards serve as the foundational content framework from the Illinois Science Assessment. Key documents are identified below:

1. A Framework for K–12 Science Education | Next Generation Science Standards
2. NGSS Front Matter
3. How to Read NGSS
4. DCI Arrangements of the Next Generation Science Standards
5. Topic Arrangements of the Next Generation Science Standards

The ISA blueprint for grades 5 and 8 is intended to provide a score to individual students that represents their performance on all dimensions of the range of NGSS standards for the elementary and middle school grade bands. The blueprint is designed to provide -- at the school or district level, depending on the number of students tested -- sub-scores that represent student performance in the domains identified below.

Science Domains:

1. Physical Science
2. Life Science
3. Earth and Space Science

Science and Engineering Practices:

1. Asking Questions and Defining Problems
2. Developing and Using Models
3. Planning and Carrying Out Investigation
4. Analyzing and Interpreting Data
5. Using Mathematical and Computational Thinking

6. Constructing Explanations and Designing Solutions
7. Engaging in Argument from Evidence
8. Obtaining, Evaluating, and Communication of Information

Table 12. Science Operational Test Blueprint for Grades 5 & 8

Domain	Section 1			Section 2			Section 3			All Sections		
	MC/TE	CR	Clusters	MC/TE	CR	Clusters	MC/TE	CR	Clusters	MC/TE	CR	Clusters
Physical	12	1	2*	6	0	1	6	0	1	24	1	4
Earth and Space	6	0	1	12	1	2*	6	0	1	24	1	4
Life	6	0	1	6	0	1	12	1	2*	24	1	4
Counts	24	1	4	24	1	4	24	1	4	72	3	12
Points	24	3	27	24	3	27	24	3	27	72	9	81

*One cluster includes six MC items; the other cluster includes six MC items and one CR item

The ISA has four performance levels, and the new names for the performance levels will mirror those for IAR: Below Proficient, Approaching Proficient, Proficient, Above Proficient.

ACT Assessment: High-School ELA, Mathematics, and Science

Beginning in spring 2025, Illinois administers the ACT to all Grade 11 students as the high school accountability assessment for ELA, Mathematics, and Science. The ACT is a standardized college admissions exam that consists of four subject tests – English, Mathematics, Reading, and Science – and an optional Writing test (essay). Illinois uses the ACT (with Writing) to measure high school achievement and college readiness in the junior year. Unlike the IAR and ISA, which are custom state assessments aligned to state standards, the ACT is a nationally administered college-readiness assessment. However, the content of the ACT is aligned to skills and knowledge Illinois students are expected to acquire by Grade 11, and it has established College Readiness Benchmarks in each subject.

Test Content: The ACT subject tests cover the following areas:

- **English:** 75 multiple-choice items in 45 minutes, assessing grammar, usage, and rhetorical skills. Passages are presented and items prompt students to demonstrate knowledge of sentence structure, grammar and mechanics, and the style and organization of the text. The content is categorized into reporting categories such as Production of Writing, Knowledge of Language, and Conventions of Standard English.

- **Mathematics:** 60 multiple-choice items in 60 minutes, covering topics from pre-algebra up through some trigonometry. Questions span content domains like Number & Quantity, Algebra, Functions, Geometry, Statistics & Probability, as well as integrating mathematical practices (referred to as “Integrating Essential Skills” and “Modeling”). A calculator is allowed on the ACT math test.
- **Reading:** 40 multiple-choice items in 35 minutes, with several passages (including one paired set of passages) on topics in prose fiction, social science, humanities, and natural science. The test assesses reading comprehension in terms of Key Ideas & Details, Craft & Structure, and Integration of Knowledge & Ideas.
- **Science:** 40 multiple-choice items in 35 minutes, measuring interpretation of data, scientific investigation, and evaluation of models and experimental results. Passages present scientific information in forms like graphs, tables, and research summaries. The content spans biology, chemistry, physics, and Earth/space sciences, though it is not a comprehensive curriculum-based test – it emphasizes scientific reasoning using provided information.
- **Writing:** 1 essay task in 40 minutes. Students must write an essay responding to a topic, demonstrating skills in Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. The ACT writing test is scored separately on a scale of 2–12 and also contributes to an ACT ELA score (which is the rounded average of the English, Reading, and Writing scores).

Scoring

Each of the four multiple-choice ACT subject tests (English, Math, Reading, Science) is scored on a scale of 1–36. The ACT Composite score is the average of the four subject test scores (also 1–36). The Writing test is scored by trained readers on a rubric and reported on a 2–12-point scale; it does not factor into the 1–36 Composite, but it is combined with the English and Reading scores to produce an ACT ELA score on the 1–36 scale (for students who take the essay). For accountability reporting, Illinois intends to use the ACT ELA, ACT Math, and ACT Science scores to categorize student performance in Grade 11 ELA, Mathematics, and Science, respectively. The ACT ELA score is essentially a combined measure of a student’s skills in English, reading comprehension, and writing. For the standard setting, the ACT ELA scores included in the impact data were calculated from students who completed all three subject tests, but for the purposes of accountability reporting, ISBE will calculate an ACT ELA score for students who did not complete the writing test by setting their writing score to zero and taking the average of students’ English, reading, and writing scores.

ACT College Readiness Benchmarks

In 2005, ACT established College Readiness Benchmarks reflecting the ACT scores of 11th and 12th grade students associated with a 50% chance of earning a B or higher grade in common first-year credit-bearing courses at a typical postsecondary institution (Allen and Scoring, 2005). The Benchmarks also correspond to an approximate 75% chance of earning a C or higher grade in these courses.

The courses selected to establish the Benchmarks were based on courses commonly taken by students in their first year of college, including English Composition for English, College Algebra for math, Social Science courses (including American History, Other History, Psychology, Sociology, Political Science, and Economics) for reading, and Biology for science. The original Benchmarks corresponded to scores of 18 on the English test, 22 in math, 21 in reading, and 24 in science. The Benchmarks were updated in the fall of 2013 (Allen, 2013) to address possible changes in college grading standards, student performance, course taking patterns of first-year college students, and alignment between secondary and postsecondary course content that transpired since the original benchmarks were established. Using a large sample of first-year students attending two- and four-year institutions, the study detected no changes in the English and Math Benchmarks (18 and 22, respectively), the Reading Benchmark increased from 21 to 22, and the Science Benchmark decreased from 24 to 23.

ACT typically uses an empirical, or Modified Briefing Book, standard setting approach (Camara et al., 2017; Haertel et al., 2012) when assisting states to establish cut scores on the ACT. This approach links ACT scores directly to academic performance in college, in contrast to content-based approaches that focus on item content and difficulty. The empirical approach establishes cut scores using a process like that used to develop the ACT College Readiness Benchmarks but is driven by the judgments of Illinois educators and other stakeholders and uses state-specific impact data as well as other relevant data included in the briefing books. This approach differs from the Briefing Book approach described by Haertel et al., (2012) in that while extensive impact and comparative data are provided in the briefing books, a pre-selected set of potential cut scores is not included. This approach is preferred for several reasons. The primary purpose of college admissions tests is to identify students who are likely to succeed in postsecondary academic environments. Such assessments are used for admissions, placement, recruitment, and talent identification because they predict GPA and grades in specific college courses. Educators in secondary schools use the tests to determine if students are on track to being college ready at the end of high school, to identify academic weaknesses that can be addressed, and to aid in postsecondary planning for students. This type of evidence prioritizes the empirical relationship between test scores and outcomes such as postsecondary enrollment, course grades, GPA, and retention.

Prior to 2025, Illinois used a separate Grade 11 assessment (SAT or ISA Science) with its own performance standards. In 2025, as part of this standard setting, Illinois is establishing cut scores on the ACT scale for Level 2 (Approaching Proficient), Level 3 (Proficient), and Level 4 (Above Proficient) in ELA, Math, and Science using a hybrid approach incorporating both content and empirical standard setting methodologies. This approach was selected rather than the empirical approach that ACT typically uses to ensure a unified approach across grade levels and assessments. The cut scores were informed not only by content expectations (PLDs for what a proficient 11th grader should know) but also by empirical data about college readiness, using the data underlying the ACT Benchmarks as one piece of evidence. Illinois’s panelists considered such data and other research when setting the cut scores, so that “Proficient” on the ACT corresponds roughly to a student being on track for college-level work. The briefing books (described in greater detail later) provided data showing relationships between performance on the ACT and multiple high school and postsecondary outcomes, which helped ground the cut score decisions in real-world outcomes.

PreACT 9 Secure and PreACT Secure

Beginning in 2025, PreACT 9 Secure and PreACT Secure are administered at Grades 9 and 10 in Illinois as preparatory, vertically aligned assessments. These tests are scaled similarly to the ACT (score ranges 1–32 for PreACT 9 Secure and 1–35 for PreACT Secure) and have similar test specifications but tailored to content appropriate for 9th and 10th graders. The PreACT Secure assessments do not have a writing test, but Illinois will use the average of the English and Reading scores as the ELA measure in grade 9 and 10. Content blueprints for the ACT suite of assessments can be found in Table 13 to Table 16.

Because PreACT 9 Secure and PreACT Secure are built to similar test specifications as the ACT, and because we sought performance levels with normative coherence across the high school grades (9–11), ACT recommended that cut scores for the PreACT Secure assessments should be statistically linked to the corresponding cut scores that were set for the ACT. Specifically, an equipercentile linking approach (Kolen & Brennan, 2004) was used to determine cut scores for grades 9 and 10 based on the Round 4 ACT cut scores. Once the ACT Approaching Proficient, Proficient, and Above Proficient cut scores were established, equipercentile linking was used to determine the scores on the grade 9 and 10 tests most closely matching the percentile ranks of the respective ACT cut scores. This ensured that the proportion of students in each performance level was approximately the same across grades 9–11. Following this step, the cut scores were reviewed and adjusted during the vertical articulation to achieve a consistent trend across all grade levels.

Table 13 to Table 16 contain the content blueprints for PreACT 9 Secure, PreACT Secure, and the ACT assessments.

Table 13. Content Blueprint for PreACT 9 Secure, PreACT Secure, and ACT English Assessments.

Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT English Assessments			
Content Blueprints	PreACT 9 Secure	PreACT Secure	ACT
Total number of operational items	33	36	75
Items per passage	10–12	11–13	14–16
Reporting Categories			
1. Production of Writing	9–11 (27–33%)	10–12 (28–33%)	22–24 (29–32%)
2. Knowledge of Language	4–6 (12–18%)	5–7 (14–19%)	11–13 (15–17%)
3. Conventions of Standard English	16–18 (48–55%)	18–20 (50–56%)	39–41 (52–55%)
Operational Passages			
1. Total Number	3	3	5
2. Genre (SSC, HUM, NSC, etc.)	0-1 each	0–1 each	0-2 each
Depth of Knowledge (DOK)*			
1. DOK 1	12–16 (36–48%)	13–17 (36–47%)	21–28 (28–37%)
2. DOK 2	7–9 (21–27%)	8–10 (22–27%)	18–24 (24–32%)
3. DOK 3	13–16 (39–48%)	14–17 (39–47%)	27–33 (36–44%)

*Webb’s Depth of Knowledge (DOK) is a framework developed by Norman Webb to categorize the complexity of educational tasks widely used in curriculum design, assessment, and instructional planning. The DOK model includes four levels, with higher levels indicating higher amounts of cognitive complexity involved.

Table 14. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Math Assessments

Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Math Assessments			
Part I: Content Blueprints	PreACT 9 Secure	PreACT Secure	ACT
Total number of operational items	30	33	60
Reporting Categories			
Preparing for Higher Math (PHM)	24 (80%)	21 (64%)	36 (60%)
Number and Quantity	3–4 (10–13%)	3–5 (9–15%)	5–7 (8–12%)
Algebra	6–8 (20–27%)	4–6 (12–18%)	7–9 (12–15%)
Functions	6–8 (20–27%)	4–6 (12–18%)	7–9 (12–15%)
Geometry	3–4 (10–13%)	3–5 (9–15%)	7–9 (12–15%)
Statistics and Probability	3–4 (10–13%)	3–5 (9–15%)	5–7 (8–12%)
Integrating Essential Skills (IES)	6 (20%)	12 (36%)	24 (40%)
Modeling	≥7 (≥23%)	≥8 (≥24%)	≥12 (≥20%)
DOK			
DOK 1 PHM	6–8 (20–27%)	6–8 (18–24%)	7–11 (12–18%)
DOK 2 PHM	11–13 (37–43%)	11–13 (33–39%)	15–21 (25–35%)
DOK 3 PHM	4–6 (13–20%)	4–6 (12–18%)	7–11 (12–18%)
DOK 1 IES	0 (0%)	0 (0%)	0 (0%)
DOK 2 IES	4–6 (13–20%)	4–6 (12–18%)	12–18 (20–30%)
DOK 3 IES	1–3 (3–10%)	1–3 (3–9%)	7–11 (12–18%)

* Some topics and skills in Preparing for Higher Math for PreACT 9 Secure as on grade level move to Integrating Essential Skills for PreACT Secure and ACT. Some topics and skills in Preparing for Higher Math for PreACT Secure as on grade level move to Integrating Essential Skills for ACT.

Table 15. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Reading Assessments

Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Reading Assessments			
Content Blueprint	PreACT 9 Secure	PreACT Secure	ACT
Total number of operational items	25	25	40
Items per passage	8–9	8–9	9–11
Reporting Categories			
1. Key Ideas and Details	13–15 (52–60%)	13–15 (52–60%)	21–24 (53–60%)
2. Craft and Structure	7–9 (28–36%)	7–9 (28–36%)	10–12 (25–30%)
3. Integration of Knowledge and Ideas	3–4 (12–16%)	3–4 (12–16%)	6–9 (15–23%)
Operational Passages			
1. Total Number	3	3	4
2. Genre			
a. Literary Narrative	1	1	1
b. Informational	2	2	3
3. Format			
a. Single	2–3	2–3	2–3
b. Pair	0–1	0–1	1
c. VQI	0	0	0–1
DOK			
1. DOK 1	3–6 (12–24%)	3–6 (12–24%)	4–12 (10–30%)
2. DOK 2	3–6 (12–24%)	3–6 (12–24%)	14–24 (35–60%)
3. DOK 3	9–16 (36–64%)	9–16 (36–64%)	10–20 (22–50%)

Table 16. Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Science Assessments

Content Blueprints for PreACT 9 Secure, PreACT Secure, and ACT Science Assessments			
Part I: Content Blueprint	PreACT 9 Secure	PreACT Secure	ACT
Total number of operational items	26	30	40
Reporting Categories			
1. Interpretation of Data (IOD)	6–13 (23–50%)	6–12 (20–40%)	16–20 (40–50%)
2. Scientific Investigation (SIN)	4–9 (15–35%)	5–12 (17–40%)	8–12 (20–30%)
3. Evaluating Models and Scientific Arguments with Evidence (EMI)	5–10 (19–38%)	6–12 (20–40%)	10–14 (25–35%)
Passage	4 operational	5 operational	6 operational
Number by Passage Format			
a. Data Representation (DR)	2	1	2
b. Research Summary (RS)	1	3	3
c. Conflicting Viewpoints (CV)	1	1	1
Number of Passages by Content Area			
a. Earth/Space Science	1	1	1–2
b. Life Science: Biology	2	2	2
c. Physical Science: Physics	0–1	1	1–2
d. Physical Science: Chemistry	0–1	1	1–2
e. Secondary Code: Engineering and Design Thinking (EDT)	1–2	1–2	0–3
Number of Items by Content Area			
a. Earth/Space Science	5–8	5–7	5–15
b. Life Science: Biology	10–14	10–13	11–15
c. Physical Science: Physics	0–8	5–7	5–15
d. Physical Science: Chemistry	0–8	5–7	5–15
DOK			
1. DOK 1	1–4 (4–15%)	1–6 (3–20%)	2–8 (5–20%)
2. DOK 2	12–20 (46–77%)	15–23 (50–77%)	20–30 (50–75%)
3. DOK 3	6–12 (23–46%)	6–12 (20–40%)	8–16 (20–40%)

Summary of Assessments

Table 17 provides a summary overview of all the assessments included in the standard setting, including content areas, grades, format, and performance level structure.

Table 17. Summary of Illinois Assessments and Standard Setting Scope

Assessment	Grades	Subjects	Format	Performance Levels (2025)
Illinois Assessment of Readiness (IAR)	3–8	ELA, Math	Online (Computer-based); some paper accommodations	4 levels (Level 1–4)
Illinois Science Assessment (ISA)	5, 8	Science	Online; scenario-based	4 levels (Level 1–4)
PreACT 9 Secure, PreACT Secure	9, 10	ELA (English & Reading), Math, Science	Online, timed (ACT platform)	4 levels (linked to ACT)
ACT	11	ELA (English, Reading, Writing), Math, Science	Online, timed (national exam)	4 levels (Level 1–4)

The ACT in Illinois is administered online in 2025, with paper available as needed.

This standard setting addressed all the above assessments in a unified event to ensure consistency of performance standards across grades. The next section describes the performance level descriptors that were used as the foundational criteria for these judgments.

Performance Level Descriptors

Clear and specific Performance Level Descriptors (PLDs) are a cornerstone of a standards-referenced standard setting process. PLDs are narrative descriptions of the knowledge, skills, and abilities that students at a given performance level are expected to demonstrate. They serve as a shared vision for panelists of what it means to be “Approaching Proficient” vs. “Proficient,” for example, and help ensure that cut score recommendations are grounded in content expectations rather than relative student performance. Prior to this standard setting, Illinois undertook a thorough update of its PLDs to align with the new four-level structure and the Illinois Learning Standards.

Development of New PLDs

As part of Illinois’s statewide effort to refresh assessment expectations and ensure vertical alignment across grade levels and subjects, the Illinois State Board of Education (ISBE) developed a comprehensive set of Performance Level Descriptors (PLDs) to serve as the foundation for standard setting. The PLDs describe the knowledge and skills students are expected to demonstrate at various levels of performance for each grade and subject area assessed by the Illinois Assessment of Readiness (IAR), Illinois Science Assessment (ISA), and the high school assessment suite. The PLDs were publicly released and are available at <https://www.isbe.net/Pages/Performance-Level-Descriptors.aspx>.

To ensure technical rigor and content validity, ISBE convened over 100 Illinois educators from across the state to serve on grade-band and subject-specific PLD development teams. These educators met twice per month from September 2024 through April 2025 to draft and refine the PLDs in collaboration with ISBE staff and content experts. Each draft was developed using a consistent framework that aligns with assessment claims and evidence statements. The resulting documents were carefully reviewed to ensure they struck an appropriate balance between clarity and rigor—referred to by ISBE as the “Goldilocks” level of specificity. The PLDs were intentionally written to reflect a wide range of student performance and were not intended to mirror the standards verbatim, but rather to serve as bridges between classroom instruction, formative assessment, and summative performance.

Structure of PLDs

The structure of the PLDs includes several distinct formats designed to support diverse stakeholders, from psychometricians and educators to parents and policymakers. Each subject and grade-level set includes high-level summary descriptors that provide an

accessible overview of student expectations at each performance level, as well as more detailed PLDs that are organized by assessment claim and aligned to the Illinois Learning Standards. These detailed PLDs describe specific student behaviors and capabilities aligned to tested content and support the interpretability of scale scores and achievement levels. In addition, many subject areas include accompanying "Samples to Success" documents, which present example test items, annotated student responses, scoring rubrics, and performance-level rationales. These exemplars are designed to illustrate how the PLDs can be operationalized through instruction and assessment.

Summary PLD

The Summary Performance Level Descriptors (PLDs) were essential for the Standard Setting meeting held in July 2025, connecting the Illinois Assessment of Readiness (IAR) and Illinois Science Assessment (ISA) to classroom instruction under the Theory of Action that academic standards reflect a collective commitment to annual student learning. These assessments measure knowledge, skills, and understanding with a common measure, revealing performance patterns to guide school and district improvement by highlighting strengths and opportunities. The PLDs define proficiency across four levels—Above Proficient, Proficient, Approaching Proficient, and Below Proficient—linking assessment data to formative assessments that shape local instruction and individual student choices. Proficiency represents a range of observable student performance characteristics, with multiple pathways allowing students to use their strengths differently within this range. A student's proficiency is shaped by the complexity of texts or stimuli, tasks, and contexts; as performance advances, students manage more challenging demands. Used in the Vertical Articulation process, the PLDs align expectations across grades within content areas, offering detailed descriptors to ensure a coherent progression that supports proficiency definition and instructional alignment.

Detailed PLD

The Detailed Performance Level Descriptors (PLDs) aligned the Illinois Assessment of Readiness (IAR) and Illinois Science Assessment (ISA) with classroom instruction and formative assessments, embodying the Theory of Action that academic standards represented a collective commitment to annual student learning. Available for English Language Arts (ELA), Mathematics, and Science across grades, these PLDs provided a common measure to reveal performance patterns, guiding school and district improvement by highlighting strengths and opportunities. More specific than summary versions, the Detailed PLDs contained information at the evidence statement or standard level, detailing performance across Below Proficient, Approaching Proficient, Proficient, and Above Proficient. In the July 2025 standard setting process, these PLDs were used to judge items, ensuring alignment with Illinois Learning Standards and precise proficiency assessment.

Standard Setting Methodology

Illinois employed a two-pronged standard setting methodology to accommodate the different natures of the assessments involved. For the Grades 3–8 IAR and ISA assessments (as well as the initial phase for the ACT), a content-based judgment method known as the Extended Modified Yes/No Angoff was used. For the ACT (Grade 11) assessments, after an initial Angoff phase, an empirical, data-informed method (Modified Briefing Book) was integrated to align the cut scores with external college readiness criteria. Both methods are well-established in standard setting literature and were tailored to Illinois’s specific context. Below we describe each in turn.

Extended Modified Yes/No Angoff Method

The Angoff method, in general, is a test-centered, item-by-item judgment approach where panelists estimate the performance of a hypothetical group of students at the borderline of a performance level. In the traditional Angoff, panelists might estimate the probability that a minimally proficient student would answer each item correctly (e.g., 0 to 100%). These probabilities are summed across items to yield a recommended cut score (the expected number of points a just-Proficient student would earn). The Yes/No Angoff is a variation that simplifies this task: instead of assigning a probability, panelists give a dichotomous judgment for each item – Yes (a borderline proficient student would likely get it correct) or No (likely get it wrong). Implicitly, this Yes/No decision is made with a certain probability criterion in mind. In our case, “likely” was defined as having at least a 1/2 chance (50%) of answering correctly. This criterion was communicated so that panelists understood “Yes” means they believe a borderline student has a reasonably high chance of success on that item.

The method is “Extended” to accommodate polytomous (multi-point) items and multiple performance levels. The IAR and ISA include items that are worth more than 1 point (e.g., an ELA prose constructed response (PCR) essay in grades 4–8 is worth 0–4 points for Written Expression and 0–3 points for Conventions; reasoning and modeling items on the IAR math assessment; CR items on the ISA science assessment; ACT writing essay etc.). The extended Angoff approach allows panelists to judge how many points a borderline student would likely score on such an item (essentially, a Yes/No for each possible point the item could earn, or a direct assignment of points). In practice, for a 4-point essay, a panelist might judge that a barely Level 3 student would likely get 2 out of 4 points (so the “judgment” for that item would be 2, which conceptually means the student would meet the response criteria for 2 points but not 3). These judgments are summed to form the cut score recommendation. The summed cuts were transformed to scaled scores, so panelists received scaled score cuts as part of their feedback.

Illinois utilized the Extended Modified Yes/No Angoff for each cut score. This means panelists went through the items multiple times – effectively once for each cut. However, to make the process efficient, we structured it in rounds where multiple cuts could be considered in tandem (see Process section for details on rounds). Each committee ultimately provided three cut scores: the score that separates Level 1 and 2 (Approaching Proficient cut), Level 2 and 3 (Proficient cut), and Level 3 and 4 (Above Proficient cut). Figure 2 below illustrates the concept of multiple cut scores on a score scale with four performance levels.

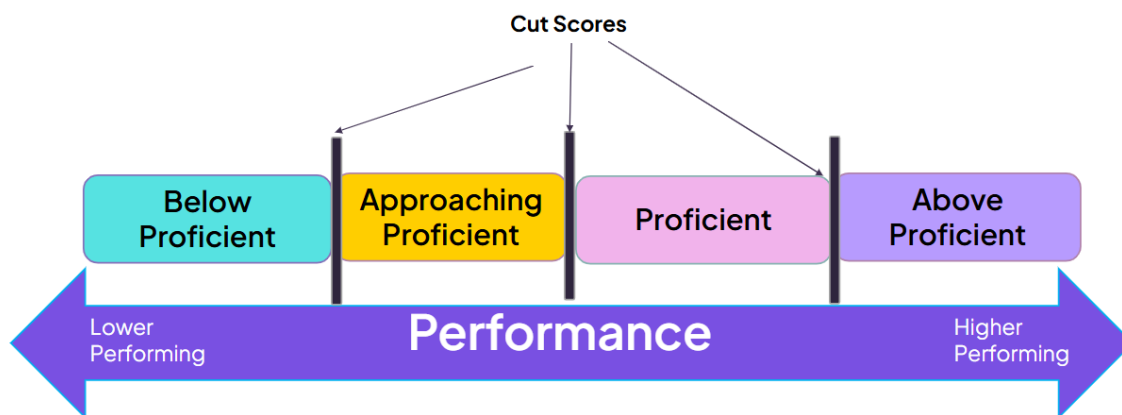


Figure 2. Performance Levels and Cut Scores

The appeal of the Yes/No Angoff is that it is relatively straightforward for educators to grasp – it mirrors their intuition of thinking of a specific student and asking “Would they get this item right?” rather than assigning abstract probabilities. It also aligns well with how tests like IAR and ISA are taken (students either answer correctly or not, or get partial credit). Research has shown that the Yes/No modification can yield results comparable to traditional Angoff while reducing cognitive load on judges (Plake & Cizek, 2012). We also chose this method because it integrates nicely with providing feedback profiles by item: for example, after a round, facilitators can show which items a panelist (or group) said “Yes” vs “No” and how that compares to actual student performance.

An important enhancement we implemented was the use of student performance profiles during the Angoff judgments. Given that IAR and ISA include essays and performance tasks that contribute significantly to the score, we provided panelists with sample student response data at various score points – essentially anonymized actual student responses or score profiles that illustrate what an actual test taker scoring at a certain level did on the test. This helps ground the panelists’ judgments in reality. For instance, before making Round 2 judgments, panelists could review a profile of a student whose total score was, say, at the panel’s tentative cut – seeing how students in that decile answered each item (which items they got right, how they performed on open-response) to judge if that seems

like the borderline performance described by the PLDs. This profile approach (sometimes called “benchmark student papers” or response maps) adds a validation layer to the Angoff judgments. It is a modification inspired by methods used in other assessments where constructed responses are significant (Davis & Moyer, 2015).

In summary, for IAR and ISA, the process was:

1. Panelists received training on the Yes/No Angoff and practiced on sample items.
2. Discussed borderline student characteristics at each cut (using PLDs and their teaching experience).
3. Executed multiple judgment rounds. Panelists made Yes/No judgments at each cut level for each item, using PLDs and item notes. After Round 1 feedback (ranges/outliers), panelists independently revised judgments in Round 2, informed by discussion, panelist feedback data, and score profiles introduced at that point. Round 3 ratings were further informed by impact data of student performance and concluded with independent judgments and final feedback. Committee medians were adopted as the panel’s recommended cut scores at each level; dispersion (range/IQR) was monitored across rounds.

This implementation preserves a direct content link between cuts and what borderline students can do.

Hybrid Angoff/Modified Briefing Book Method (ACT – Round 3 & 4)

For the ACT standard setting, Illinois employed a hybrid methodology that merged the content-based Angoff judgments with an empirical, data-driven approach often referred to as a “Modified Briefing Book” method. The rationale for this hybrid approach is that the ACT is a nationally norm-referenced test with well-researched external performance indicators, so incorporating that information can help ensure the Illinois performance standards are meaningful in terms of college readiness.

The process for ACT consisted of four rounds in two phases. Phase 1 mirrored the process described earlier for the IAR and ISA, and Phase 2 transitioned into an empirical approach described below. In order to integrate the two phases instead of having an abrupt transition between the two approaches, ACT introduced a subset of the empirical data between Rounds 1 and 2, including impact data (the percentages of students scoring within each performance level and at or above each ACT score point) and the probabilities of earning an A, B or higher, or C or higher grade in first year college courses associated with each ACT score point. This ensured a more integrated approach in which the Round 2 cut scores were based on content and informed by outcome data.

Phase 1 (Rounds 1–2): Modified Angoff (Yes/No) – This mirrored the process described earlier for the IAR and ISA. In Round 1, panelists on each ACT subject committee for the multiple choice tests (English, Math, Reading, and Science) made Yes/No judgments for borderline performance at each cut level on each item. Panelists on the ACT Writing committee made judgments about the number of points that students at the borderline of each performance level would be expected to earn in each domain of the scoring rubric. Because ACT had only one grade (Grade 11) for standard setting, these committees were only setting cuts for one test each, but multiple performance levels. In Round 2, they reviewed feedback, including score profiles and impact data. After an initial discussion of the Round 1 results, panelists received an introduction to the ACT Benchmark data and use of ACT scores in Illinois colleges and universities and had an additional opportunity to discuss the new information before making their Round 2 judgments. By the end of Round 2, each subject committee had an initial set of cut score recommendations for that subject test based on item-level judgments.

Phase 2 (Rounds 3–4): Data-Driven Adjustment (Modified Briefing Book) – After Round 2, the process transitioned to an empirical approach. Panelists were given Briefing Books specific to their subject area (math, science, or ELA) containing a wealth of data about ACT scores and external criteria, and substantial time was dedicated to explaining the contents of the Briefing Books page by page. The Briefing Books included descriptive, impact, and outcomes data for several ACT-tested populations. The focal population was the Illinois ACT-tested juniors in spring of 2025. Descriptive and impact data were also provided for Illinois students in grades 9 and 10 who took PreACT 9 Secure and PreACT Secure in the spring of 2025, since these data would be used in the vertical articulation on Friday. Additional data populations included aggregated data from eight states (Alabama, Arizona, Hawaii, Montana, Nebraska, Nevada, Wisconsin, and Wyoming) who administered the ACT with writing to 11th grade students statewide in 2023, 2024, and 2025. Descriptive information included student demographics (gender and race/ethnicity), average ACT Composite and subject test scores, and the percentages of students meeting each of the ACT College Readiness Benchmarks for each population. Impact data included the percentages of students scoring at or above each score point for each subject test. The probabilities of achieving a grade of A, B or higher, or C or higher in specific college courses as a function of ACT score were provided in both tables and graphs.

Table 18 is an example of the probabilities of success and impact data for math, and Figure 3 is an example of the probabilities of success showing graphically how the probability of obtaining an A, B or higher, or C or higher grade in College Algebra increases with higher ACT Math scores. Historical performance of Illinois students on the IAR, ISA, and SAT in 2023 and 2024, and the most recent NAEP results for Illinois and the nation were provided to provide context. Post-secondary impact data, including college enrollment rates and long-term college outcomes, were also provided. College enrollment rates by ACT score were provided for the ACT-tested graduating classes of 2022, 2023, and 2024 for two populations: national graduates and graduates from other census-tested states. Illinois-specific enrollment rates were not provided because low percentages of students in Illinois took the ACT in recent years. Long-term college outcomes were based on a study by ACT that followed a national sample of nearly 200,000 students who enrolled in college through six years of post-secondary education (Noble and Radunzel, 2012). Four outcomes were provided by ACT section test score: second year retention for students enrolled at a two-year college, second year retention for students enrolled at a four-year college, six-year bachelor's degree completion rates for students enrolled at a four-year college, and six-year associate or bachelor's degree completion rates for students enrolled at a two- or four-year college.

Table 18. Probability of Success in College Algebra and Percentage of Students At or Above Each ACT Math Score

ACT Math Score	Probability of Success			Percentage At/Above				
	A prob	B or higher prob	C or higher prob	IL Grade 11 ACT	Census State Juniors	Census State Juniors	Census State Juniors	Census State Grads
				2025	2025	2024	2023	2024
36	0.89	0.94	0.96	1	0	0	0	0
35	0.86	0.93	0.95	3	1	1	1	1
34	0.83	0.92	0.95	4	2	1	1	1
33	0.78	0.90	0.94	5	2	2	1	1
32	0.74	0.88	0.92	5	2	2	2	2
31	0.70	0.86	0.91	6	3	3	3	3
30	0.64	0.83	0.90	7	4	3	3	3
29	0.58	0.80	0.89	8	4	4	4	4
28	0.51	0.77	0.87	10	6	5	5	6
27	0.45	0.73	0.85	12	7	8	7	8
26	0.39	0.69	0.83	15	10	10	10	11
25	0.34	0.64	0.80	19	14	13	13	14
24	0.29	0.59	0.78	23	16	16	16	17
23	0.23	0.55	0.75	26	19	19	19	20
22	0.20	0.51	0.73	29	22	22	22	23
21	0.16	0.46	0.70	32	25	24	25	26
20	0.13	0.40	0.66	35	28	28	28	29
19	0.11	0.35	0.63	39	32	32	32	33
18	0.09	0.30	0.60	45	39	38	38	39
17	0.07	0.26	0.56	52	46	47	47	48
16	0.05	0.22	0.51	63	59	60	61	62
15	0.04	0.19	0.46	75	73	74	75	76
14	0.03	0.16	0.43	84	85	87	87	88
13	0.02	0.13	0.39	91	92	93	93	94
12	0.02	0.11	0.35	95	96	96	97	97
N-count	70,461			139,151	285,969	282,998	273,839	287,401

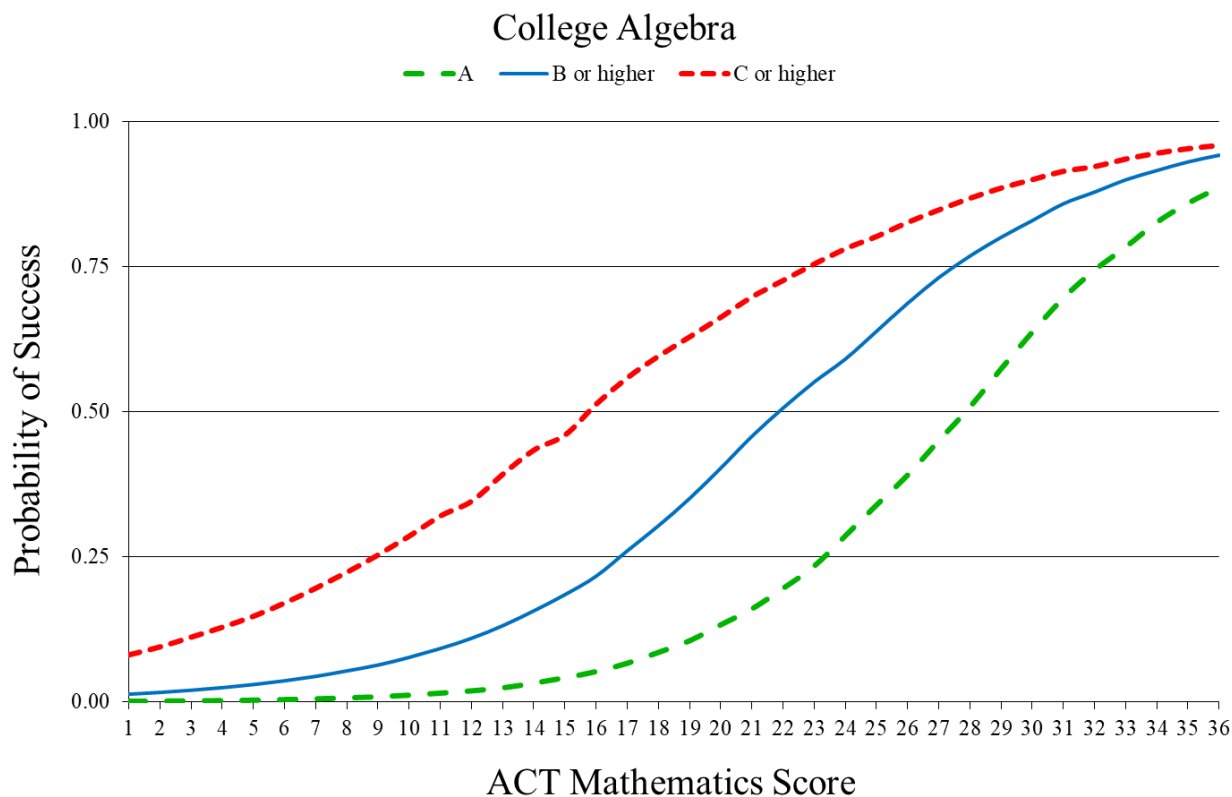


Figure 3. Plot of Probabilities of Success in College Algebra by ACT Math Score

Armed with these data, the panelists then reconvened, but importantly reconfigured. A subset of panelists from each of the English, Reading, Writing committees combined to form a single ELA committee, meaning that the five subject-specific committees combined into three broader committees for ELA, Math, and Science, and the focus turned from item-level performance to score-level performance. The ELA committee focused on determining cut scores on the ACT ELA score rather than the individual subject test scores. The ACT ELA score (1–36) is an average of the English, Reading, and Writing subject tests rather than a single administered test. Panelists were shown how the Round 2 cut scores on the English, Reading, and Writing tests combined into ELA cut scores. The Math and Science committees were largely comprised of the same panelists as in previous rounds, with an additional panelist joining the Science committee to provide a workforce perspective for Rounds 3 and 4.

Each of these committees reviewed the Round 2 recommended cuts (from Angoff) and compared them to the empirical indicators. Panelists were instructed to compare the Round 2 cuts to the data in the briefing books and consider the additional information, whether they found it compelling, and which information was more important or influential in making their cut score determinations. There is no one “right answer” – it’s a policy decision how rigorous to set the cut scores – but the data ensures the decision is informed.

Panelists were instructed to consider both the content judgments *and* the empirical data in making final recommendations.

For Rounds 3 and 4, the judgment task focused on scores and probabilities rather than item-level judgments. Panelists were asked to think about minimally Proficient students in their subject area, and think about their likely success in a first-year, credit-bearing college course in that subject area (e.g., College Algebra, Biology, English Composition, or Social Science courses) in terms of what grade would they be likely to earn (A? B or higher? C or higher?) and the probability that they would earn that grade (e.g., 50%? 60%?). Then they would select the ACT score that best reflects what they see as their probability of achieving that grade. This process would be repeated for Approaching Proficient and Above proficient, with each panelist providing a total of three judgments for each round. Panelists were instructed to provide their individual judgments, as consensus was not required.

After Round 3, there was additional opportunity for discussion, as well as a direct comparison of the Round 2 and Round 3 results to discuss any differences emerging from the two different methodologies. Panelists also reviewed the results across subject areas in a general session, with an opportunity for panelists to share their thoughts across the three committees. After this discussion, panelists provided their final Round 4 cut score recommendations, using the same procedure described above for Round 3.

Throughout Phase 2, facilitators guided discussions and answered any questions to ensure panelists understood the data correctly. It's easy to be overwhelmed by numbers, so the Briefing Book was structured in a digestible way and considerable time was dedicated to presenting and describing the contents of the Briefing Books, page by page. For example, panelists were walked through several examples of the probabilities of success in the tables and figures so they could see that at ACT Math score X, the probability of a C or higher in College Algebra is Y%. They considered, "If we set Proficient at X, is that aligned with our expectation of a proficient 11th grader?"

The hybrid approach draws from research by Camara et al. (2017) on combining expert judgment with empirical student outcome data to set cut scores that are meaningful beyond the test itself. Haertel et al. (2012) describe the "Briefing Book" method where panelists are given relevant external data ("briefing materials") to inform standard setting in later rounds. ACT's Modified Briefing Book approach differs from the approach described by Haertel et al. (2012) in that ACT provides briefing books containing extensive impact and comparative data, but does not provide a pre-selected set of potential cut scores. Instead, ACT provides data for the full range of scores from which panelists select cut scores for each performance level. In Illinois's case, this approach was especially pertinent because one of the goals was to ensure that "Proficient" in Grade 11 signifies a level of preparation consistent with college readiness standards (of which the ACT Benchmarks are indicators).

By merging the Angoff and Briefing Book methods, Illinois leveraged the strengths of both:

1. The Angoff rounds ensured the cut scores were rooted in the content of the test and what Illinois educators expect their students to know by each level.
2. The Briefing Book rounds ensured that those expectations were calibrated against real-world outcomes and comparisons (preventing, for example, setting a cut that is unrealistically high or low in terms of college readiness).

The unified standard setting methodology was reviewed with the Illinois Technical Advisory Committee (TAC) prior to implementation and the Modified Angoff methodology was piloted in an internal dry run (conducted in April 2025 on sample data for Grade 3 ELA and Math as a proof of concept). The TAC and ISBE approved the approach, noting that it was a innovative yet appropriate strategy given Illinois’s assessment transition.

In execution, the panelists found the process to be logical. Qualitatively, some panelists from the ACT committees commented that they found the hybrid approach beneficial. The combination of content and data gave them confidence that their final recommendations would make sense to stakeholders. Meanwhile, for the IAR/ISA committees, sticking with the Angoff-only method was appropriate because those tests don’t have similar external benchmarks.

In conclusion, Illinois’s standard setting methodology was a comprehensive, multi-method approach. It adhered to best practices in standard setting – combining expert judgment, empirical validation, and vertical articulation – to produce cut scores that are content-valid, aligned with educational goals, and empirically grounded in outcomes beyond K–12. The next sections describe the participants who carried out this methodology and how the meeting was prepared and executed.

Participants

The success of a standard setting study depends on the quality and representativeness of its participants. Illinois and its assessment partners (Pearson and ACT) undertook careful recruitment and selection to ensure that the standard setting panels consisted of knowledgeable, diverse, and dedicated individuals. In addition to panelists, several other roles including facilitators, psychometricians, content experts, and observers were involved in the process.

Panelist Recruitment and Composition

Recruitment Process

ISBE issued a call for standard setting panelists in early 2025, inviting educators and stakeholders across Illinois to apply or be nominated. Over 900 applications were received, reflecting strong interest from the field. From this pool, 155 panelists were selected to serve on committees spanning Grades 3 through 11. Selection criteria included geographic region, school district size and type (urban, suburban, rural), role, content expertise, years of experience, and student populations served (e.g., English Learners, students with disabilities, gifted). Priority was given to current or recently practicing educators at the relevant grade levels.

Committee Assignments

A total of 19 committees were convened: 14 for the IAR/ISA assessments (ELA Grades 3–8, Math Grades 3–8, and Science Grades 5 & 8) and 5 for the ACT assessments (English, Reading, Writing, Mathematics, Science). Each committee had approximately 12 panelists, though some (e.g., Grade 4 ELA and Math) were intentionally larger to facilitate vertical overlap across grades. Most committees met for two days, with some panelists serving on multiple grade-level groups to strengthen continuity. In total, 147 unique panelists filled 254 committee “slots.”

Geographic and Institutional Representation

Panelists represented all seven ISBE regions, 118 unique districts or schools, and 35 counties across the state. Table 15 provides a summary of regional representation and institutional diversity.

Table 19. Geographic and Institutional Representation

Variable	Summary
Geographic Location	Area 1 - (North/Northeast): 41.5% Area 7 - (Chicago/Inner Suburbs): 12.9% Area 3 - (West Central): 12.2% Area 2 - (Northwest): 10.9% Area 5 - (Southwest): 10.2% Area 6 - (Southeast): 6.8% Area 4 - (East Central): 5.4%
Representation	~120 unique districts, schools, and educational organizations
Counties Represented	45+ unique counties

Panelist Demographics

Efforts were made to ensure diversity in panelist demographics and experiences. The majority of panelists were female, with substantial representation from educators of color and those serving high-need districts.

Table 20. Panelist Demographics

Variable	Category / Sub-Category	Percentage
Gender	Female	83.0%
	Male	16.3%
	Prefer not to answer / Missing	0.7%
Race/Ethnicity	Caucasian / White	57.8%
	Black or African American	17.7%
	Hispanic or Latino	11.6%
	Asian	4.1%
	Two or more races	1.4%
	Middle Eastern or North African	2.7%
	Other	0.7%
	Prefer not to answer / Missing	4.1%

Professional Experience

Panelists brought substantial experience in K–12 education, with many holding advanced degrees. On average, panelists had more than 18 years of teaching experience and over 20 years in the field of education.

Table 21. Professional Experience (N = 145 reporting)

	Mean	Median	Min	Q1	Q3	Max
Total years of teaching experience	17.8	17	2	11.5	23.5	44
Years in current position	7.5	5	1	2.5	10	32
Total years in education	20.4	20	2	15	25	50

Specialized Roles and Experience

Panelists represented a diverse array of roles—including classroom teachers, instructional coaches, curriculum leads, and administrators—with extensive experience supporting special populations and varied instructional contexts. Notably, eligibility for participation on the standard setting panel required panelists to have classroom experience within the previous two years. This criterion was deemed essential to ensure that panelists brought current, practical insight into instructional practices, curriculum standards, and the nuances of student learning. Given the content-heavy nature of the standard setting process, which relied on thorough familiarity with standards, evidence statements, and indicators of student mastery, panelists actively engaged in teaching were best equipped to interpret and apply these expectations. Classroom teachers, in particular, remain closest to the daily realities of student performance and implementation of standards, making their contributions invaluable to the committee’s work.

Table 22. Specialized Roles and Experience

Experience Category	Percent Reporting Experience
Experience in classroom (last two years)	100.0%
Honors or Talented/Gifted Experience	26.1%
IEPs or 504 Plans Experience	76.0%
English Language Learners Experience	42.6%
Instructional Coach Experience	17.9%
Curriculum Lead Experience	42.0%
School Administrator Experience	8.2%
Other Special Population Experience	9.3%
4-Year Higher Ed Experience	5.9%
2-Year Higher Ed Experience	11.6%
Career or Technical College Experience	2.9%
Workforce Development Experience	6.0%

Panel Composition by Committee

The 19 committees were designed to balance representation across grades and subjects, while also promoting vertical alignment. Table 19 summarizes the number of panelists and representative roles for each committee. Table 24 shows how vertical articulation was built into the design of the committee structure. Most panelists that served on committees on Monday and Tuesday served on committees for an adjacent grade that met on Wednesday and Thursday. The panelists that served on an 8th grade committee and an ACT committee missed the first day of training for the ACT assessments. However, the training was identical to the training that they had already received for the general session and standard setting methodology training. They did need to time to review the ACT assessment that they were assigned to along with the PLDs. This step was accomplished by having a working lunch with on Tuesday and a working breakfast on Wednesday in the committee room so that they could review those materials with the ACT facilitator.

Table 23. Committee Composition

Panel	Subject/Grade	Panelists	Roles Represented
1	ELA – Grade 4	18	Teachers, instructional coaches, administrators
2	ELA – Grade 6	11	Teachers, coaches, curriculum coordinators
3	ELA – Grade 8	12	Teachers, literacy director
4	Math – Grade 4	18	Teachers, specialists, interventionists, coaches
5	Math – Grade 6	11	Teachers, coordinators, curriculum directors
6	Math – Grade 8	12	Teachers, specialists, coaches, district leadership
7	Science – Grade 8	12	Teachers, STEM, special education, administrator
8	ELA – Grades 3	12	Teachers, coordinator, resource teacher
9	ELA – Grades 5	12	Teachers, specialists, coaches
10	ELA – Grades 7	12	Teachers, literacy director, specialists
11	Math – Grades 3	12	Teachers, interventionists, special education, coaches
12	Math – Grades 5	11	Teachers, coordinators
13	Math – Grades 7	12	Teachers, coaches, curriculum specialists
14	Science – Grades 5	11	Teachers, special education, instructional leader
15	ACT – English	9	HS teachers, department chairs, curriculum directors
16	ACT – Reading	11	HS teachers, department chairs, administrators
17	ACT – Writing	10	HS teachers, directors, department chairs
18	ACT – Math	14	HS teachers, coordinators, special education

Panel	Subject/Grade	Panelists	Roles Represented
19	ACT – Science	11	HS teachers, administrators, instructional coaches

Table 24. Design of Committees

Committee	Mon	Tues	Wed	Thurs.	Fri. Vertical Articulation
ELA3			ELA3		14 members
ELA4	ELA4				
ELA5			ELA5		
ELA6	ELA6				
ELA7			ELA7		
ELA8	ELA8				
ACT (English, Reading, & Writing)		ACTELA			
MAT3			MAT3		
MAT4	MAT4				
MAT5			MAT5		
MAT6	MAT6				
MAT7			MAT7		
MAT8	MAT8				
ACT (MATH)		ACTMATH			
SCI5			SCI5		6 members
SCI8	SCI8				
ACT (SCI)		ACTSCI			

All panelists were compensated for their time and travel (if not local to Springfield). Their accommodations and meals during the meeting were arranged by Pearson’s event staff. Many panelists expressed that the opportunity to participate in such an important policy-setting activity was an honor and a professional growth experience. Indeed, Illinois sought not just representation but also that panelists would become ambassadors for the new standards, helping explain and support them when they returned to their districts.

Facilitators and Support Staff

Each committee was led by a facilitator. Facilitators are crucial to standard setting – they guide the panelists through the agenda, ensure the methodology is properly executed, and keep discussions productive and on schedule. For this meeting: – IAR and ISA Committees: There were 7 facilitators (one per IAR/ISA committee room) provided by ACS Ventures (a consulting firm with expertise in standard setting). These facilitators had prior experience running standard setting panels, and they were trained specifically on Illinois’s process and

materials. They were also versed in the Illinois Learning Standards to field content questions. One ACS Ventures lead facilitator coordinated across committees to ensure consistency.

Table 25. Lead Roles and Responsibilities for Pearson and ACT Staff

Name	Title/Department/Role
TAC Member	
Will Lorie, Ph.D.	Senior Associate, National Center for the Improvement of Educational Assessment (NCIEA); TAC member and independent observer of Standard Setting
ISBE Team Members	
Rae Clementz	Exec Director of Data, Accountability & Assessment
Angela Foxall	Director of Assessment • Assessment
Laura Quimby	Supervisor • Assessment
Shu-Ren Chang, Ph.D.	Principal Consultant • Assessment
Shuwan Chiu, Ph.D.	Principal Consultant • Assessment
Heather Colwell	Principal Consultant • Assessment
Megan Forness	Principal Consultant • Assessment
Pam Hartwig	Principal Consultant • Assessment
Victoria Henderson	Principal Consultant • Assessment
John Hicks	Principal Consultant • Assessment
Sarah Leonard	Principal Consultant • Assessment
Robert Potempa	Principal Consultant • Assessment
Milton Leathers	Director of Accountability • Accountability
Lindsay Record	Interim Director • Public Relations
Dusty Rhodes	Principal Consultant • Public Relations
Januari Trader	Principal Consultant • Public Relations
Meagan Budke	Principal Consultant • Standards and Instruction
Kimberly Johnson	Principal Consultant • Standards and Instruction
Christy Bryan	Project Manager • Information Technology
Trevor Cottle	Principal Consultant • Multilingual/Language Development
Pearson Team Members	
Tracy Gardner, Ph.D.	Senior Measurement Advisor, SS Lead
Yong Luo, Ph.D.	Lead Research Scientist, Lead of Psychometric Work
Erik Iwersen	Test Development Manager
Gulsah Gurkan-Kilicoglu	Primary Research Scientist
Mary Allen	Program Manager
Angie Bauer Wisner	Project Manager, Psychometrics
Lise Scheuermann	Math Content Lead
Tiffany Hartmann	ELA Content Lead
Christopher Altermatt	Science Content Lead
7 IAR & ISA Facilitators	Facilitate each committee - ACS Ventures
4 IAR & ISA Stat Analysts	Run the data for the committees
Bailey True	Statistical Analyst
Emma Zimmerman	Statistical Analyst
ACT Team Members	
Joann Moore, Ph.D.	Lead Research Scientist, SS Lead for ACT, Facilitator for ACT Science
Abby Montgomery	Director of Program Management
Patty Ferrel	Lead Program Manager
Shalini Kapoor, Ph.D.	Facilitator, ACT English & ELA
Ann Arthur, Ph.D.	Facilitator, ACT Math
Nola Daley, Ph.D.	Facilitator, ACT Reading

Name	Title/Department/Role
Kristen Phipps	Facilitator, ACT Writing
Stacy Dreyer	ACT English Content SME
Mariah Steele, Ph.D.	ACT Reading Content SME
Roxanne Swim	ACT Writing Content SME
Scott Johanningmeier	ACT Math Content SME
Jay Thomas	ACT Science Content SME
ACS Ventures Team Members	
Chad Buckendahl, Ph.D.	Founding Partner
Sarah Alahmadi, Ph.D.	Psychometrician
Melia Franklin, Ed.D.	Senior Associate
Russ Keglovits	Assessment Specialist
Kyle Killworth	Psychometric Associate
Deborah Schnipke, Ph.D.	Senior Psychometrician
Kelley Stethen	Psychometric Associate

ACT Committees

ACT provided 5 facilitators for the five subject committees in Phase 1, and then 3 lead facilitators for the combined ELA/Math/Science discussions in Phase 2. One of ACT's research scientists served as the lead facilitator overseeing the ACT standard setting (ensuring the Angoff procedure and data review were done correctly across subjects). These facilitators had deep knowledge of ACT content and college readiness research (ACT's standard setting team includes research scientists and psychometricians, some of whom were involved in developing ACT's Benchmarks and linking studies).

Facilitators

Facilitators underwent training prior to the meeting. Pearson conducted two training sessions:

1. An initial training (via webinar) to familiarize facilitators with the general standard setting process, the online portal (Moodle) usage, the test designs, and their scripts/instructions.
2. A second training that took a deeper dive into the standard setting process and focused on the feedback data, impact data, and the score profiles.

During the meeting, facilitators were responsible for:

1. Keeping time and making sure the committee stayed on the agenda (with flexibility if needed for discussions).
2. Presenting content (via PowerPoint) during training segments, such as going over sample items and modeling how to make a Yes/No judgment.

3. Moderating discussions after each round of ratings, ensuring everyone who wanted to speak had a chance, and that the conversation stayed focused (for example, avoiding tangents about unrelated policy issues).
4. Ensuring that no undue influence occurred – e.g., if an observer or a particularly vocal panelist tried to dominate, the facilitator would intervene to keep it balanced.
5. Checking that panelists were not confused about their tasks; if any confusion arose, addressing it immediately (possibly one-on-one during breaks if needed).
6. Communicating with the analysis team to know when feedback data was ready to present, etc.

Additionally, Pearson had a Standard Setting Lead (facilitator) on site overseeing the entire event. The Pearson lead (along with the ACT lead, as needed) floated between rooms, answered questions, and ensured consistency in procedure. They would, for instance, confirm that each facilitator was adhering to the script (to ensure uniform instructions across committees).

Standard Setting Roles. Other staff included:

1. **Data Analysts & Psychometricians:** Pearson and ACT each had a team of psychometricians on site or on call. Their job was to quickly compile the ratings after each round, compute the feedback statistics (like median cut scores, impact data) and produce reports/graphs for facilitators to share with panelists. They double-checked calculations to avoid errors. Given the tight turnaround (often less than 20 minutes during a break to crunch data), this team was well-prepared with pre-written code and templates.
2. **Content Specialists:** A few content experts from ISBE, Pearson, and ACT were present to answer content-specific questions.
3. **Observers:** ISBE had staff and leadership observing the process in each room (typically 1–2 per room). Additional observers included some members of the Technical Advisory Committee and Pearson/ACS staff who were not facilitating. Observers were instructed not to interfere, but they could answer *policy* questions if. Observers provided a link to policy without influencing judgment. As noted, the number of observers was kept small (≤ 3 in a room) to avoid intimidating panelists.
4. **Logistics and Technical Support:** Pearson’s event management team managed the venue, registration, etc. A tech support person was on site to handle any issues with technology (for example, if a panelist’s laptop froze while viewing items, or if Wi-Fi issues arose). The environment was secure with password protected Wi-Fi; all panelists had unique logins for the standard setting site, and no one outside the room could access the content.

General Meeting Setup. The standard setting took place in a conference center in Springfield. Rooms were arranged with round tables for panelists to sit in small groups (approx. 4 per table), and a projector/screen at the front for the facilitator. Panelists at each table were mixed by background to encourage diverse discussion. This table grouping was mainly for discussion purposes; all ratings were individual. Name placards, notepads, and printed reference materials (like hard copies of PLDs or the agenda) were provided at each seat.

The collaborative yet controlled environment helped panelists feel comfortable contributing. Many noted in evaluations that the facilitators and staff created an atmosphere of professionalism and respect, where their judgments were valued. This positive environment is reflected in the evaluation survey results, where participants strongly agreed that they had sufficient opportunity to express their views and felt their expertise was respected (see Evaluation section).

In summary, the participants in the Illinois standard setting included a diverse panel of educators and other experts as panelists, guided by skilled facilitators, and supported by a team ensuring everything from data crunching to technical setup ran smoothly. This robust human infrastructure was essential for the complex 5-day process to be executed with fidelity.

Preparation

Extensive preparation was required to conduct a week-long standard setting event for multiple assessments. In the months leading up to July 2025, Pearson and ACT coordinated with ISBE to develop materials, set up the standard setting website, pilot the process, and handle countless logistical details. This section outlines the key preparatory steps that were completed to ensure the meeting's success.

Planning and Coordination

A formal Standard Setting Plan was drafted and iteratively refined between August 2024 and April 2025. The plan detailed the methodology, committee structure, schedule, and resource needs. It was reviewed by ISBE and the state's Technical Advisory Committee (TAC) in Fall 2024 and Winter 2025. Feedback from TAC was incorporated as requested by ISBE. By spring 2025, a final plan was approved by ISBE, which served as the blueprint for execution.

Given that both Pearson and ACT were involved as partners, close coordination was essential. Regular meetings (bi-weekly then weekly as the event neared) were held among Pearson's standard setting lead, ACT's standard setting lead, and ISBE's assessment team to monitor progress on deliverables and to synchronize tasks.

A crucial step was the Internal Dry Run. On April 4, 2025, ISBE and Pearson conducted a focused dry run to evaluate two content-based standard-setting methodologies in response to TAC questions about clarity, efficiency, and the staging of impact data. The session functioned as a miniature rehearsal using Grade 3 ELA and Grade 3 Mathematics forms, with ISBE leadership and staff serving as panelists and ACT observing to ensure methodological consistency for the high-school assessments. The primary purpose was to test whether either approach better simplified the judgment task for panelists while keeping the process firmly grounded in content before any consideration of impact or external data.

Panelists received concise training on each method, completed one round of ratings, and then engaged in a facilitated discussion to reflect on the experience. To limit order effects, method order was counter-balanced by content area. Consistent with TAC guidance, the review of impact data and any complementary empirical indicators occurred only after content judgments had been made and discussed.

Across the day, we gathered structured feedback on ease of use, confidence in ratings, time and practicality, and overall method preference. For the Extended Modified-Angoff approach, we also prepared and reviewed illustrative performance profiles to support

interpretation after the rating round. Impact summaries were generated at the end of each method sequence to compare the distribution of students across performance levels under the two approaches.

The pilot findings were encouraging. Impact data were broadly similar across methods for both Grade 3 ELA and Mathematics, suggesting either approach can yield comparable classification outcomes when implemented with clear training and facilitation. However, panelists expressed a clear preference for the Extended Modified–Angoff process—especially for ELA—citing the straightforward yes/no judgment prompt, score profiles, and a generally lighter cognitive load. They also noted that very few items could be classified into a category lower than Proficient for the Modified ID matching method, which raised concerns for the panelists.

Findings and materials from the dry run were reviewed with TAC members on April 9, 2025. Based on the results and panelist feedback, the team proceeded with the planned methodologies, implementing minor refinements to the instructions, materials, and PLDs to further streamline execution and enhance clarity in the materials and directions.

Standard Setting Website and Materials

Pearson developed a dedicated standard setting website for the event, utilizing the Moodle learning management system. Each committee (grade/content) had its own course on the site, accessible only to its panelists and facilitator. Table 26 includes the list of materials that was included for each standard setting committee. The site was designed to be the central hub for panelists during the meeting:

1. It housed all the reference materials (PLDs, agendas, instructions, etc.) in digital form.
2. It provided links for panelists to launch TestNav to “experience the assessment” – i.e., take or review the test form.
3. It hosted the judgment surveys – custom online forms where panelists entered their Yes/No judgments for each item (or directly input cut scores in Round 3). These were essentially interactive questionnaires that recorded each panelist’s ratings.
4. It included feedback data pages – after each round, once data was ready, facilitators could enable a page that showed each panelist their own cut score and anonymous statistics (like a histogram of cuts, impact data charts, etc.).
5. It also included the evaluation survey at the end.

Table 26. Panelist Materials – IAR & ISA

Material	Online	Paper
A. Agendas	Y	Y
B. Test Design Documents	Y	N
Math High Level Blueprint	Y	N
Math Test Blueprint Math (one per grade for grade 3-8)	Y	N
Math Evidence Statements (one per grade for grades 3-8)	Y	N
Math Structure	Y	N
ELA High Level Blueprint	Y	N
ELA Test Blueprint (one per grade for grade 3-8)	Y	N
ELA Evidence Statements (one per grade for grades 3-8)	Y	N
ELA Literary Task Models	Y	N
Science High Level Blueprint	Y	N
Science Test Blueprint (one per grade for grade 5 and 8)	Y	N
C. Rubrics, Anchor Papers, & Scoring Expectations		
Math Rubrics, Anchor Papers, Math Test Scoring Expectations	Y	N
ELA Rubrics, Anchor Papers, PCR Writing Guide for Teachers & Students	Y	N
Science Rubrics, Anchor Papers	Y	N
D. Performance Level Descriptors		
Performance level descriptors (PLDs): Policy – 1 set for all	Y	N
Performance level descriptors (PLDs): Summary – 14 sets	Y	N
Performance level descriptors (PLDs): Detailed – 14 sets	Y	Y
Performance level descriptors (PLDs): Samples to Success (to be reviewed during Thursday session as time permits by volunteers) – 14 sets	Y	N
PLD comment form (for ISBE use only) – The PLDs can not be changed.	N	Y
E. Test Forms Info.		
“Experience the Assessment” items: same form as SS Form (2025 Online 1) – 14 sets	Y	N
“Experience the Assessment” answer key: same form as SS Form (2025 Online 1) – 14	Y	N
“Experience the Assessment” notes form – 14	N	Y
Item comments form (all tests)	N	Y
Item map (all tests)	Y	Y
F. Practice Judgments		
Math Practice judgment items (only for grades 4, 6, 8): Taken from 2025 Online2	Y	N
Math Practice judgment answer key (grades 4, 6, 8): Taken from 2025 Online2	Y	N
ELA Practice judgment items (only needed for grades 4, 6, 8): 2025 Online2, Unit 1	Y	N
ELA Practice judgment answer key (only needed for grades 4, 6, 8): 2025 Online2, Unit 1	Y	N
Science Practice judgment items (grades 5 & 8): Taken from ISBE practice items online.	Y	N
Science Practice judgment answer key (grades 5 & 8): ISBE practice items online.	Y	N
Practice judgment survey (all tests) – 14	Y	N
G. Quizzes & Surveys		
Rounds 1-3 judgment surveys	Y	N
Rounds 1-3 Readiness quiz	Y	N

Material	Online	Paper
Evaluation surveys (Appendix E)	Y	N
H. Score Profiles (per test) – 14 total for Math, ELA, Science	Y	N
I. Rounds 1–3 Feedback data	Y	N

Pearson’s tech team pre-populated the site with all necessary content. Before the meeting, User Acceptance Testing (UAT) was conducted by Pearson staff and ISBE staff to verify that each site worked properly. The website was made secure such that panelists could only access their committee’s content, and only during the meeting when the facilitator opened access (Moodle allows sections to be hidden until needed, so panelists were unable to jump ahead).

ACT created a Google Drive site to store non-secure documents for panelists to access during the standard setting event. A separate folder was created for each panel: English, Math, Reading, Science, Writing, and ELA. Within each folder were the PLDs (policy, summary, and detailed) for high school for that subject area, the ACT test blueprints, the ACT College Readiness Standards (which were provided for reference but not explicitly used during the standard setting), and a list of links to the quizzes and forms that each panel would access during the standard setting.

Microsoft Forms was used to collect data from panelists throughout the standard setting event, including the readiness surveys, rounds of judgments, and evaluations. The Forms were linked to Excel spreadsheets using Power Automate, which allowed data to automatically flow to OneDrive where facilitators could access and download the data for analysis and presentation. The R Statistical Software Shiny package (v4.2.1; R Core Team, 2022; Chang et al., 2022) was used to tabulate results and create dashboards showing the results of each round of data collection.

Panelists received two folders of materials during the standard setting event: a materials folder and a briefing book. The materials folders contained confidential restricted information and were closely monitored throughout the event. These materials included an ACT test form that was used for the Experience the Assessment activity and Round 1 and 2 judgments; an item map containing the item numbers, passage numbers, reporting categories, PLD claims associated with each item, and answer keys; the ACT practice items and associated item maps; printed copies of the PLDs for that subject area; and for the writing test, two sets of test prompts and annotated anchor papers (one for experiencing the assessment and the Rounds 1 and 2 judgments, and one for the practice round), and the writing scoring rubric.

Briefing books were provided to panelists after the Round 2 discussion. The briefing books contained impact evidence by subject area and ACT test score, as well as descriptive information about the impact data populations, and comparative impact for the IAR, ISA,

and SAT state assessments and NAEP. The ACT impact data populations were based on several ACT-tested student populations of interest. The focal data population was the Illinois statewide-tested juniors of the spring of 2025. Descriptive and impact data were also provided for Illinois students in grades 9 and 10 who took PreACT 9 Secure and PreACT secure in the spring of 2025, since these data would be used in the vertical articulation on Friday. Additional data populations included aggregated data from eight states who administered the ACT with writing to 11th grade students statewide in 2023, 2024, and 2025, and the performance of the 2024 ACT-tested high school graduates from the eight census-tested states. The 2025 ACT-tested juniors in Illinois were the population of primary interest since these are the most recent results and closest approximation to the population that will be impacted by the new cut scores¹. Multiple years of data from other census state juniors were provided to illustrate how impact data can fluctuate from year to year due to differences in the student cohorts, and to help panelists understand that they should not overly focus on a specific percentage. Performance and demographics of the census state juniors were similar across 2023, 2024, and 2025, with some small year-to-year fluctuations.

Performance data from the 2024 ACT-tested high school graduates from the census-tested states was also provided as a comparison group, since these results are released publicly by ACT each year and panelists are likely familiar with this group. Whereas the census junior populations contain students' 11th grade state test scores, the grad class populations contain students' most recent ACT scores as of the July following their senior year of high school. The 2023 census state juniors can be compared to the 2024 census state grads as it is the same cohort of students, and while we expect to see some growth from 11th to 12th grade, 84% of the 2024 census state grads most recently took the ACT during their junior year, so the increases in average ACT scores from the 2023 juniors to the 2024 graduates do not reflect growth across the entire cohort.² The aggregated census states included Alabama, Arizona, Hawaii, Montana, Nebraska, Nevada, Wisconsin, and Wyoming and are presented to provide comparative information from other states administering the ACT (with writing) to virtually all 11th grade students. Comparing students in Illinois to the other census-tested states, Illinois students had higher performance than students in the other census-tested states.

¹ The data that ACT presented during the standard setting for the IL Census Juniors may not be an exact match to the final data results compiled by ISBE due to potential differences in data cleaning and quality checks. ACT's data sample included all students who had an ACT score for at least one subject test.

² Lower average ACT ELA and writing scores were seen when comparing the 2023 census state juniors to the 2024 census state grads. This is because students who take the ACT on a national test date can choose to not take the writing test, and therefore would not earn an ELA score. Eighty-four percent of students in the 2024 census state grads took writing compared to 97% of the 2023 census juniors. Differences in performance of students who chose to test with or without writing are likely why we see this score difference.

The briefing books also contained impact data from the IAR, ISA, and SAT in 2023 and 2024. The percentages of students scoring at or above each performance level were provided from 2023 and 2024 for math, science, and ELA. The most recent NAEP impact data was also provided, including the percentages of students scoring at or above each NAEP performance level in Illinois and the nation in grade 8, and national only in grade 12, in math, science, and reading.

Post-secondary impact data, including college enrollment rates and long-term college outcomes, were also provided. College enrollment rates by ACT score were provided for the ACT-tested graduating classes of 2022, 2023, and 2024 for two populations: national graduates and graduates from other census-tested states. Illinois-specific enrollment rates were not provided because low percentages of students in Illinois took the ACT in recent years. Long-term college outcomes were based on a study by ACT that followed a national sample of nearly 200,000 students who enrolled in college through six years of post-secondary education (Noble and Radunzel, 2012). Four outcomes were provided by ACT section test score: second year retention for students enrolled at a two-year college, second year retention for students enrolled at a four-year college, six-year bachelor's degree completion rates for students enrolled at a four-year college, and six-year associate or bachelor's degree completion rates for students enrolled at a two- or four-year college. The briefing books for each subject can be found in the Appendix.

Materials Development Process

Extensive materials were required for the successful implementation of the standard setting meeting, and careful attention was given to their development, review, and approval. Adhering to this process was critical given the volume and complexity of materials needed to support the work, and the steps were intentionally built into the overall Standard Setting schedule. For the IAR and ISA, all standard setting materials and PowerPoint presentations were developed by the Pearson Standard Setting Lead, but were subject to review and approval by ISBE. The SS Lead first developed template materials and then sought feedback from internal Pearson stakeholders, including the Program Team, Content Team, Performance Scoring Center (PSC), and ACT. Based on this feedback, the SS Lead revised the template materials and submitted them to ISBE for review. After receiving and incorporating ISBE's feedback, the SS Lead once again engaged the Program and Content Teams to review the revisions. Finalized template materials were then submitted for ISBE sign-off, followed by a final quality control check prior to use in the meeting.

For ACT assessments, a parallel process was followed. The ACT Standard Setting lead developed all standard setting materials in collaboration with Pearson, with ISBE also responsible for review and approval. Materials were first drafted and circulated among ACT stakeholders for feedback, after which revisions were made. ISBE then reviewed the updated materials, and additional revisions were incorporated based on their input. ACT

stakeholders were once again engaged for feedback before the materials were finalized. As with the Pearson-led process, the materials were submitted to ISBE for sign-off, and a final quality control check was conducted to ensure accuracy and readiness.

Facilitator Guides and Slides

Detailed PowerPoint presentations were developed for:

1. The General Session (used at the kickoff to orient everyone to the purpose and the big picture).
2. Each Breakout Session (specific to content area and round, guiding panelists step-by-step through each phase of the process). For consistency, slide templates were standardized across committees, with facilitators reading from notes so that all panelists got the same information in the same words.
3. The Vertical Articulation session (slides to facilitate cross-grade discussion and data review on Day 5).

These slides were reviewed by ISBE in advance and refined in multiple rounds. Slides that introduced impact data were carefully worded to avoid implying a “right” percentage of students that should be proficient, as that decision is ultimately judgment-based.

Practice Items and Training Materials

To train panelists on the Angoff method, Pearson selected a set of practice items that were similar in format and content to the operational tests. Practice items for ELA and Math were drawn from the other online version of the 2025 operational test. For Science, the sample items were from the paper version of the released items on the ISBE website. We selected the items on the paper form because the 2025 operational assessment had technology enhanced items as field test items, but not as operational items yet. For the ACT committees, ACT provided 10 sample items for each subject test; for Writing, panelists looked at annotated anchor papers and scores to calibrate their understanding of writing performance. These practice materials were compiled into a Practice Judgment Exercise for Day 1, so panelists could go through the motions on something low-stakes before judging real items.

A “Readiness Quiz” was developed to verify that panelists understood the instructions and concepts. After the training and practice, but before Round 1, panelists were prompted to answer a short quiz on the website.

Data Preparation

In advance, Pearson and ACT prepared all the data that would be needed for feedback and the briefing books. This included writing code to compute:

1. For each round's cuts: the % of students in each performance level (impact data) given 2025 actual score distributions.
2. For each item: classical item stats (conditional p-values and item means) showing performance by student deciles, so that panelists could review conditional p-values by decile for single point items and conditional item means for multi-point items for groups of students across the score scale.
3. The ACT Briefing Book data: ACT created printed booklets for each subject area containing the impact and comparative data described above. Much of the information was also incorporated into dashboards to show the data in the briefing books linked to each round of cut scores.

Vertical Articulation Tool

Pearson created a simple spreadsheet tool that could take the draft cut scores from all grades and visually display them on one chart. It also allowed trying out adjustments and immediately seeing the effect on impact. This was used on Day 5 for vertical articulation and content coherence discussions, allowing panelists to explore questions such as, "What happens to the impact data if we raise the Grade 4 math Proficient cut by 2 scaled score points?"

All data sets were tested ahead of time using preliminary 2025 data (as dummy) to ensure the programs ran correctly. Once 2025 full student data was available in late June, final data files were generated and double-checked.

Meeting Logistics

The standard setting event took place in Springfield, IL from July 14–18, 2025. Pearson secured a venue with sufficient breakout rooms. There were at least 10 breakout rooms (to run committees in parallel) plus a large room for general sessions and ACT combined sessions, as documented by the room layout plan in the Standard Setting Plan.

Daily agendas were created (see Appendix A). Each committee had a specific schedule of activities. For instance, on Monday–Tuesday, the committees for Grades 4, 6, 8 were running; ACT was executed on Tuesday–Thursday; Grades 3, 5, 7 ran on Wednesday–Thursday, etc., as described earlier.

Security measures

All test content used (items, answer keys, student responses) were handled as secure materials. Panelists could not keep any notes that contained specific item information – all such notes were collected and shredded at the end. Devices were monitored; panelists

used computers provided by Pearson to access the Moodle site or the ACT Google Drive site, which was on a secure network. Observers and staff had to sign non-disclosure agreements as well. The importance of maintaining confidentiality of test items and cut score results until official release was stressed at the start of the meeting (and repeated in slides).

Contingency planning

The preparation team also had backup plans in case of unexpected issues. For example, if a facilitator became ill, an alternate facilitator was on standby. We also had remote data analysis support.

Panelists later commented in evaluations that the meeting was well-organized and that materials were adequate and clear, reflecting the extensive preparation (i.e., survey results show high ratings for organization and clarity – see Evaluation section).

IAR & ISA Process

The standard setting meeting itself spanned five days and followed a carefully structured agenda. See Table 23 for a high-level summary of the five day agenda. This section provides a narrative of how the process unfolded, including the General Session, breakout sessions for the individual committees (with training and multiple judgment rounds), the integration of the ACT hybrid process, and the Vertical Articulation on the final day. This table focuses on the processes used for IAR and ISA. The ACT process is described later in this report.

Table 27. Meeting Agenda Overview: Modified Angoff Yes/No Approach

Day	Activities
One week prior	<ul style="list-style-type: none"> Pre-work
Days 1 & 3: Monday July 14, 2025 Wednesday, July 16, 2025	<ul style="list-style-type: none"> General session (Day 1 only): Welcome, Assessment Overview, SS Overview Experience the Assessment PLD review Develop Borderline Descriptors Standard Setting Training (Day 1 only) Practice judgments activity and discussion (Day 1 only)
Days 2 & 4: Tuesday, July 15, 2025 Thursday, July 17, 2025	<ul style="list-style-type: none"> Readiness Survey Round 1 judgments and feedback/discussion Round 2 judgments and feedback/discussion Round 3 judgments and feedback/discussion Process Survey Final Evaluation

Day	Activities
Day 5: Friday, July 18, 2025	<ul style="list-style-type: none"> • Introductions • Purpose of Vertical Articulation • Cross Grade PLD Review • Cross Grade Impact Data Review • Recommended Changes to Round 3 Results • Cross Content Area Review • Recommended Changes to Round 1 Articulation Results • Evaluation

General Session

On Monday, July 14, 2025, at 8:00 a.m., the standard setting kicked off with a General Session attended by all panelists who were beginning on Day 1 (this included those assigned to the Monday-Tuesday committees: ELA 4,6,8; Math 4,6,8; Science 8. ACT held a similar session on Tuesday for all high school groups across 5 committees (English, Reading, Writing, Mathematics, & Science). A general briefing was conducted on Wednesday morning for the second wave of panelists who started that day. Due to the structure of the committees, this was only required for Grade 5 Science. The general session welcomed participants, introduced key personnel, provided a big-picture overview of the assessment designs and the standard-setting process, and reviewed historical context before content breakouts.

During the general session, an ISBE representative greeted the group and thanked them for contributing their time and expertise to this important process. ISBE emphasized the significance of the outcome – these cut scores would define how Illinois reports student achievement and would impact students, schools, and the public’s understanding of educational progress. This helped underscore the responsibility of the task and motivate panelists.

Next, Pearson’s lead facilitator provided an Orientation:

1. An overview of the assessments involved (brief recap that we are setting standards for IAR 3–8, ISA 5 & 8, and ACT, aligning everything to the new performance levels).
2. A restatement of the goal: to recommend cut scores that correspond to the performance level definitions and PLDs.
3. A note on process integrity: panelists were reminded that there are no pre-determined “right answers” for cut scores; their collective professional judgment is what determines the recommendations. They were encouraged to be honest and thorough.

4. Explanation of confidentiality and security: All participants signed NDAs at check-in and were reminded not to discuss specific test items or emerging results outside the meeting. Phones were silenced and stored during sessions.

The general session also introduced the performance level policy definitions (the broad descriptions of Below Proficient, Approaching Proficient, Proficient, and Above Proficient that apply across grades).

A high-level explanation of the standard setting method was given without diving into details. It was noted that Illinois was using a method involving multiple rounds of judgments and that some of them would involve looking at data. The presenter mentioned that after all groups finish, a vertical review would promote coherence.

Crucially, the general session tried to build a sense of community and purpose. Panelists from different committees could see the scale of the effort – nearly 19 committees all working in concert. This helped them feel part of a larger mission. It was highlighted that no one individual’s perspective or one committee’s outcome stands alone. They were informed that together they were helping build a continuum of expectations from elementary to high school. This was motivating to panelists and underscored why vertical articulation would happen on Day 5.

By 9:15 a.m., the general session concluded and panelists dispersed to their assigned breakout rooms to begin the detailed work.

Breakout Sessions – Training and Round 1

Once in their committee breakout rooms, each group met their facilitator and fellow panelists. The facilitator began with introductions – each panelist often introduced themselves (name, school/district, role, experience). This served to break the ice and also let everyone appreciate the diversity in the room.

They then went through an orientation to materials and the online system. Each panelist logged into the Moodle site on a laptop. The facilitator walked them through the layout: where to find the PLDs, where to click to launch the test, how to submit judgments. A quick systems check ensured everyone’s login worked and they could navigate.

The committee then proceeded through the structured training activities:

Experience the Assessment

Panelists were first asked to take the test from a student’s perspective. Using the TestNav link on the site, they opened the actual spring 2025 form of the test they were setting standards for. Panelists either actively answered the items or just read through all items and answers to familiarize themselves with the content and difficulty. This step is important

because it grounds panelists in the actual test experience. It took roughly the same length as the student testing time (for example, about 75–90 minutes for an ELA test including reading passages). If any panelists finished early, they were free to take a short break while others finished. Facilitators encouraged them to note any items they found particularly easy or hard for students, any surprises in content, etc., to discuss later.

After all were done, a short debrief occurred: the facilitator asked, “How did you find the test? What are your initial impressions of overall difficulty or breadth?” Panelists shared observations. This was kept brief and general, as detailed item discussion would happen later if needed. But it helped validate that everyone saw what the test entailed.

Performance Level Descriptor (PLD) Review

Next, the facilitator shifted focus to the Performance Level Descriptors (PLDs), which outlined grade-level expectations for the subject and grade. Hard copies of the detailed versions of the PLDs were distributed to each panelist. The facilitator prompted discussion among panelists to ensure everyone understood the distinctions between levels.

A key emphasis during this review was the concept of “borderline” performance. For each level, panelists were encouraged to consider students whose scores placed them at the lower end of a performance level—the threshold, such as those who were just barely classified as Proficient. While the PLD describes the full scope of a performance level, these borderline students likely would not consistently demonstrate all aspects of the descriptor, but would be considered to be right on the lower edge of the level.

When necessary, the facilitator asked panelists to identify and discuss the differences between adjacent PLDs, such as distinguishing what shifts from Approaching to Proficient. This process helped panelists calibrate their mental models for student performance, especially at the borderline between Proficient and students scoring in the middle or higher ranges of the Proficient level.

Borderline Student Definition & Discussion

The PLDs were used to guide a discussion of borderline performance for each performance level. A borderline student was located at the cut score, or a student with a score close to the lower boundary of a performance level (e.g., the student who just reached the Proficient performance level). To define the expectations for the demonstrated performance of the borderline student, the panelists worked in their table groups to consider the key distinguishing characteristics of students' performance at the borderline of each level. Panelists were encouraged to jot down key words to help them conceptualize borderline performance and how it differs from typical performance. The panelists were then led through a whole-group discussion to review key considerations from the small groups, with

the opportunity to provide feedback and summarize their ideas across the levels for cohesiveness. The panelists started by discussing borderline performance for the Proficient level to establish a baseline expectation, followed by discussion of Above Proficient and Approaching Proficient. They considered the following types of questions while discussing and defining borderline performance:

- If a student demonstrated just enough performance to get into a performance level, how was that performance described? What did they demonstrate with respect to these skills?
- What differentiated a student with performance at the borderline of the performance level relative to a student in the middle or upper end of this level?
- What differentiated a student with performance at the borderline of the performance level relative to the upper end of the lower performance level?
- What are the key distinguishing characteristics of “just barely” performance compared to performance at other points on the continuum of performance for that level?

Methodology Training

The panelists were trained on the steps followed to make their recommendations using the Extended Modified Yes/No Angoff method, including an orientation to each judgment process component and how they were used, such as the standard setting website, item maps, judgment record forms, and judgment surveys. Significant time was spent discussing the judgment question that panelists were asked to consider for each task, as outlined in Table 28 “Considering a variety of students at the lower end of the 'Proficient' category, would most students get this item correct?”

Table 28. Judgment Question Breakdown

Question Component	Description
Considering a variety of students	Students are so diverse in their knowledge, skills, and abilities, so it is hard to consider just one student. Therefore, we ask panelists to consider a variety of students so that they will consider a diverse set of students and a diverse set of pathways to arrive at a particular result.
at the lower end of the [specific]* proficiency level	The panelists should reference the borderline descriptions for the performance level to determine the knowledge and skills a student at the borderline would be expected to demonstrate.
*Insert Proficient, Above Proficient, or Approaching Proficient	These questions must be asked for each proficiency level cut: <i>Proficient, Above Proficient, and Approaching Proficient</i>
would	When considering expected student performance on an item, the panelists need to consider how a student <i>would</i> achieve rather than how they <i>should</i> achieve to reflect realistic expectations of student performance on the assessment.
most students	Panelists need to consider whether they believe at least 50% of the borderline students with knowledge and skills at the borderline would answer the item correctly.
get this item correct?	The final judgment for each item is a “yes” or “no” judgment.

To determine whether most of the borderline students would answer an item correctly, panelists were asked to: (a) determine the knowledge and skills needed to provide a correct response to the item, (b) review the knowledge and skills a borderline student would demonstrate, and (c) estimate the likelihood (probability) that a borderline student would provide a correct response. If half, or 50%, of the borderline students for a particular performance level would likely provide a correct response, the answer to the judgment question for the task at that performance level was “yes.” If not, the answer was “no” for the item. The training focused on the following points:

- The performance level and its PLDs were thought of as representing a range of performance on the performance scale.
- The panelists’ job was to decide what the lower borderline of that range described by the PLDs was.
- Panelists thought of the lower borderline in terms of a student who was “minimally qualified” to be in the performance level.
- Panelists used the PLDs and the concept of borderline performance to individually decide what “minimally qualified” meant and whether or not most students would get that item correct (if rating a single-point item)

For polytomously scored items, the question became: “Considering a variety of students, which score point most likely represented the most common response for students at the lower end of this proficiency level?” The panelists reviewed the scoring rubric and anchor

papers to help better understand how students responded. The specific questions were summarized in Table 29. These questions were repeated for each cut.

Table 29. Generic & Specific Questions for Extended Modified-Angoff (Yes/No) Question by Item Type

Item Type	Question
Dichotomously scored: Generic	Considering a variety of students at the lower end of this proficiency category, would most students get this item correct?
Dichotomously scored: Specific for each Proficiency Level	<p>Proficient: Considering a variety of students at the lower end of Proficient, would most students get this item correct?</p> <p>Above Proficient: Considering a variety of students at the lower end of Above Proficient, would most students get this item correct?</p> <p>Approaching Proficient: Considering a variety of students at the lower end of Approaching Proficient, would most students get this item correct?</p>
Polytomously scored: Generic	Considering a variety of students, which score point, most likely represents the most common response for students at the lower end of this proficiency level?
Polytomously scored: Specific for each Proficiency level	<p>Proficient: Considering a variety of students, which score point, most likely represents the most common response for students at the lower end of Proficient?</p> <p>Above Proficient: Considering a variety of students, which score point, most likely represents the most common response for students at the lower end of Above Proficient?</p> <p>Approaching Proficient: Considering a variety of students, which score point, most likely represents the most common response for students at the lower end of Approaching Proficient?</p>

Practice Judgment Activity

At the end of the training session, panelists made practice judgments prior to beginning the actual judgment rounds to get a feel for the range of items and student responses they encountered during the judgment task, gain experience reviewing and making judgments for each item, and build their confidence in their understanding of the task they were asked to complete. A set of practice items had been selected by Pearson and approved by ISBE that included a range of different item types and item difficulty. The practice items for ELA and Math had come from the 2025 Online 2 form. Due to the structure of the committees, practice items were needed for grades 4, 6, and 8 for ELA and Math. ELA had used Unit 1 from the 2025 Online 2 form. Math had used a miniature set of items across all subclaims A-D from 2025 Online 2 for this practice set. Science required practice items for both grades 5 and 8. Those practice items had been derived from the Science practice items on the ISBE website for the paper forms. For science, the paper forms were selected because the 2025 science forms did not have TEI operational items.

The goal of the practice activity was to ensure panelists understood and followed the correct process to make judgments and submit their online ratings successfully. The

facilitator walked through the judgment materials for the first two items to ensure panelists knew where to locate key information for making their judgments. There was no overlap in these practice items with items on the standard setting forms, so the group was encouraged to engage in discussion to ensure understanding of the method and how to use the PLDs, anchor papers, and other materials to guide ratings, etc. After the panelists finished making their practice judgments, the facilitators discussed the results, highlighting notable variation in the panelist cut score agreement and provided the panelists an opportunity to briefly share their perspective regarding their cut score selection.

Readiness Quiz

Before making judgments for each round, panelists indicated their readiness to participate in the standard setting activity and confirmed their understanding of the judgment task by responding to the questions in the readiness quiz. The Rounds 2 and 3 readiness quiz also included a question about the panelists' understanding of the feedback data, as shown in Figure 4. Panelists were not able to start the judgment task for a round until they answered "yes" to each question.

Are you ready to begin the judgment activity?

Select one:

Yes

No

Before starting the activity, select a response for each of the following questions.

Do you understand the feedback data provided?

Select one:

Yes

No

Do you understand your task for the judgment activity?

Select one:

Yes

No

Figure 4. Readiness Quiz

Judgment Rounds for IAR/ISA Committees

For all three rounds of ratings, panelists made judgments for all the items on one complete test form. The items were presented in the same order as they were presented to students taking the assessment. When reviewing each item, panelists started with the Approaching Proficient borderline descriptions and compared the borderline PLDs with the knowledge, skills, and abilities required by the item. The panelists then made judgments for the same item for the rest of the performance levels (e.g., Proficient and Above Proficient) before proceeding to the next item. Judgments were recorded in the Pearson Standard Setting site, using the Item Judgment Survey. Once the panelists had completed making judgments for all items, they submitted their judgments for analysis.

After all panelists had completed the judgment activity, data analysts collected the item judgments from each participant from the site, performed the necessary analysis of the data (with appropriate quality control checks of the data and results), and created feedback data that was provided to panelists.

The participants engaged in this judgment activity for three rounds. After the first and second rounds, panelists were provided with feedback on their current recommendations, as well as information on the recommendations of others in their group and relevant information from actual student performance on the assessment. During each feedback discussion, panelists were encouraged to listen to other panelists and consider the rationales given for their judgments but not to feel pressured to change their judgment to reach consensus.

The cut score recommendations were presented on the scaled score metric. For illustrative purposes, an example of panelist feedback is provided over the following pages. Table 30 shows an example of the Round 1 cut score recommendations. Table 31 shows an example of the Panelist Judgments. Table 32 shows an example of the conditional p-values. Here is an example of the script that the facilitator discussed with the panelists on how to use the conditional p-values:

“This chart displays the average performance of students for each item, based on scaled score ranges divided into 10 deciles, d1 to d10, with each decile representing approximately 10% of students. The columns (d1 to d10) represent these deciles, with d1 at the left (lowest-scoring 10% of students) and d10 at the right (highest-scoring 10%). The rows represent all items on the assessment. A subset of sample items is shown, including CR and MC, which are on the same scale, making them comparable.

This data is an example of what you will review for the assessment that you are working on and can help you understand how student performance across score ranges aligns with our performance levels. The color coding, based on the legend, indicates the performance of students within each decile for the item. To make the interpretation

more user friendly, we coded the p-value, or percent correct for single point items and percentage of points earned for multi-point items into categories. The green (↑) represents p-values values ≥ 55 , yellow (↔) for values < 55 and ≥ 45 , and red (↓) for values < 45 meaning less than 45% of students in that decile category go that item correct, helping us assess expectations for threshold students at Approaching Proficient (APP), Proficient (PRO), and Above Proficient (ABP) levels.

For example, the first item in the first row shows consistent strong performance across all deciles, especially for higher-scoring students, indicated by green (↑) coloring. In contrast, the last two items displayed on this slide (highlighted in red) indicates challenging performance, particularly for lower-scoring students (APP range) and middle scoring students, with red (↓) coloring indicating p-values < 45 , though the highest scoring students (ABP) performed better with green (↑) in the highest decile. Other items show mixed patterns, appearing challenging for students in lower deciles, but manageable for students in the higher deciles. This data can provide additional insights into how students performed at various points along the score scale.

We share this after Round 1, not before, so you can focus first on the item content, borderline descriptors, and performance level descriptors during your initial ratings, ensuring your judgments are rooted in qualitative aspects before exploring data. Now, this profile becomes a valuable tool. Items with a lot of green (↑) across deciles (like the first item) are easy for most students, while items with a lot of red (↓) are difficult. Mixed patterns with yellow (↔) and green (↑) are moderate. Use this info to guide your Round 2 reflections.”

Table 30. Round 1 Cut Score Recommendations: Science

	N	Mean	Median	Min	Max	Q1	Q3
APP Raw Score	18	29.61	27.00	13.00	51.00	22.00	38.00
PRO Raw Score	18	53.06	55.00	26.00	67.00	46.00	60.00
ABP Raw Score	18	69.61	71.00	53.00	78.00	67.00	72.00

Table 31. Round 1 Sample Panelist Judgments

SeqNo	UIN	APP	PRO	ABP
1_MC		No	No	No
2_MC		No	No	No
3_MC		Yes	Yes	Yes
4_MC		No	Yes	Yes
5_MC		No	No	Yes
6_MC		No	No	Yes
7_MC		Yes	Yes	Yes
8_MC		No	No	Yes
9_MC		No	Yes	Yes
10_MC		No	Yes	Yes
11_MC		No	No	Yes
12_MC		No	Yes	Yes
13_MC		No	No	No
14_MC		No	No	Yes
15_MC		Yes	Yes	Yes
16_MC		Yes	Yes	Yes
17_MC		No	No	No
18_MC		Yes	Yes	Yes
19_MC		Yes	Yes	Yes
20_MC		No	Yes	Yes
21_MC		No	No	Yes
22_MC		No	Yes	Yes
23_MC		Yes	Yes	Yes
24_MC		Yes	Yes	Yes
25_CR		2	3	3

Table 32. Round 1 Sample Profile of Conditional P-Values

Ability	Max	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
		650-680	681-699	700-709	710-716	717-723	724-731	732-741	742-756	757-773	774-850
	1	→	↑	↑	↑	↑	↑	↑	↑	↑	↑
	1	→	↑	↑	↑	↑	↑	↑	↑	↑	↑
	6	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	1	↓	↓	↓	→	↑	↑	↑	↑	↑	↑
	1	↓	→	↑	↑	↑	↑	↑	↑	↑	↑
	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
	1	↓	↓	→	→	→	↑	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	→	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	1	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	2	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
	3	↓	↓	↓	↓	↓	↓	↓	↓	→	↑
	1	↓	↓	↓	↓	↓	↓	↓	→	↑	↑
	2	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
	1	↓	↓	↓	↓	→	→	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑

A final set of feedback that panelists received was the Impact data, but they did not receive that until after Round 2. Figure 5. Round 3 Impact Data (Example): ELA shows an example of Impact data for ELA. It should be noted that committees only saw Impact data for the the grade level that they were working on at this stage.

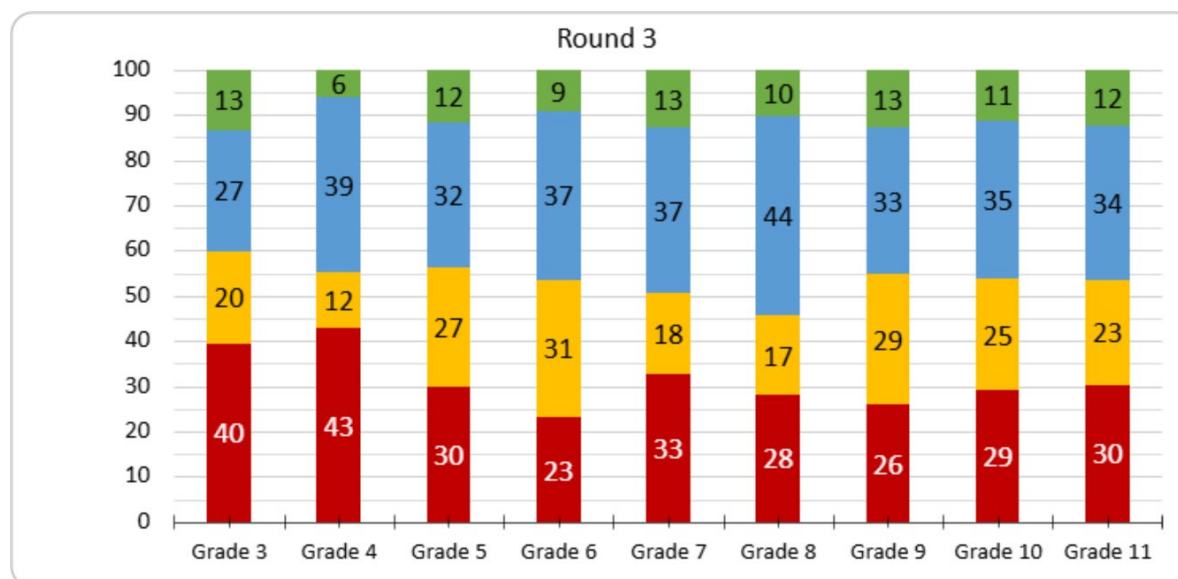


Figure 5. Round 3 Impact Data (Example): ELA

Evaluations

The standard setting panelists completed two structured evaluation surveys during the Illinois Assessment of Readiness (IAR), Illinois Science Assessment (ISA), and ACT Standard Setting Meetings, along with a separate survey for Vertical Articulation participants. These evaluations served two primary purposes: to provide ISBE and Pearson with actionable feedback for improving future standard setting meetings, and to serve as important validity evidence supporting the current process. All surveys were administered electronically. Results are presented in Appendices D–G.

Evaluation Survey #1 was administered after the practice activity early in the meeting. Its purpose was to gather feedback on the usefulness of training activities, the development of borderline descriptions, and the adequacy of time allocated for training and practice. This survey was completed by panelists across all grade-level committees in ELA, Math, and Science. Results showed that the training activities were consistently rated as “Useful” to “Very Useful,” with particular value placed on experiencing the actual assessment, discussing item scoring, and reviewing the Performance Level Descriptors (PLDs). Activities such as the practice judgments and borderline description discussions were also rated highly in terms of usefulness in assisting recommendations. Across both the Day 1 committees (Grades 4, 6, and 8) and the Day 3 committees (Grades 3, 5, and 7), panelists indicated that the amount of time devoted to training and practice was generally “Adequate” to “More Than Adequate.” Overall, Evaluation #1 results confirmed that the training structure and content effectively prepared panelists for the judgment tasks. See Appendix D (for grades 3–8) and F (for ACT).

Evaluation Survey #2 was administered after Round 3 of the standard setting process. This survey focused on panelists’ perceptions of the judgment process, the usefulness of feedback data, their confidence in both borderline descriptions and final cut score recommendations, the adequacy of facilities and tools, and the quality of interactions among participants. Panelists across committees for Grades 4, 6, and 8 (Day 2) and Grades 3, 5, and 7 (Day 4) reported that the judgment process and feedback data were generally “Successful” to “Very Successful,” with impact data and student score profiles consistently viewed as especially valuable. Confidence in both the borderline descriptions and the final cut score recommendations was rated from “Somewhat Confident” to “Very Confident,” with most panelists leaning toward the higher end of this scale. The meeting facilities, online tools, and materials were generally rated “Adequate” to “More Than Adequate,” with computers and the standard setting website receiving particularly strong ratings. Panelists also reported that they were able to express their opinions, ask questions, and interact with fellow panelists effectively, and most felt their perspectives were respected by facilitators and peers. See Appendix E for grades 3–8 and Appendix F for ACT.

Finally, a separate survey was administered to participants in the Vertical Articulation meeting. This survey collected feedback on the usefulness of the stakeholder discussion and the clarity of the recommendations that emerged. Results again pointed to positive perceptions of the process and outcomes. See Appendix G.

The evaluation results demonstrate that panelists and stakeholders viewed the training, process, and tools as effective and appropriate. The surveys confirmed that panelists had confidence in their work and that the structures in place supported meaningful engagement. These findings provide evidence of the quality of the meeting and the validity of the cut score recommendations produced.

ACT Standard Setting Process (Concurrent Rounds)

Meanwhile, in parallel on Tuesday–Thursday, the ACT committees followed their hybrid process, which we outline in this section. The high-level agenda is provided below followed by a broad summary of each round.

Table 33. Meeting Agenda Overview: Briefing Book Method - ACT

Day	Activities
Day 2: Tuesday, July 15, 2025	<ul style="list-style-type: none"> • General session: Welcome, Assessment Overview, SS Overview • Experience the Assessment • PLD review • Develop Borderline Descriptors • Standard Setting Training • Practice judgments activity and discussion
Day 3: Wednesday, July 16, 2025	<ul style="list-style-type: none"> • Review/Debrief, welcome panelists from grade 8 standard setting • Readiness Survey • Round 1 judgments (Angoff Y/N) and feedback/discussion • Overview of Empirical Methodology • Overview of ACT Benchmarks data and Probabilities of Success • Discussion • Round 2 judgments (Angoff Y/N) and feedback/discussion
Day 4: Thursday, July 17, 2025	<ul style="list-style-type: none"> • Review/Debrief • Introduce Data Books • Training & Round 3 judgments (Empirical) • Discussion • Coherence across Subjects and Methodologies • Round 4 judgments (Empirical) • Panelists Complete Process Evaluation • Present Final Recommendations

Day 1 (Tuesday): The ACT subject committees (English, Math, Reading, Science, Writing) began in a general session then moved to their own breakout rooms. The general session largely mirrored the Monday presentation for the IAR and ISA. Once in breakout rooms, panelists completed the “*Experience the Assessment*” activity by taking their section of the ACT test. Panelists received printed test forms for their subject area and marked their answers in the test booklet for the multiple choice test; for writing, panelists used Notepad on their laptops to complete the essay, and for math, panelists were allowed

to use the Desmos calculator (www.desmos.com) on their laptop. To underscore the high-stakes nature of the ACT, panelists completed their section of the test under standard timed conditions and the facilitators read the verbal instructions that students receive when they take the ACT. After experiencing the assessment, panelists reviewed the item maps including the answer keys, reporting categories, and PLD claims measured by each item. They then reviewed the detailed PLDs for their subject area – grades 11–12 for ELA and grades 9–12 for math and science. Following the same procedures as those for the IAR and ISA, they engaged in a discussion of borderline performance, followed by training for the Modified Angoff Yes/No standard setting methodology, a practice round of judgments, and an evaluation at the end of the day.

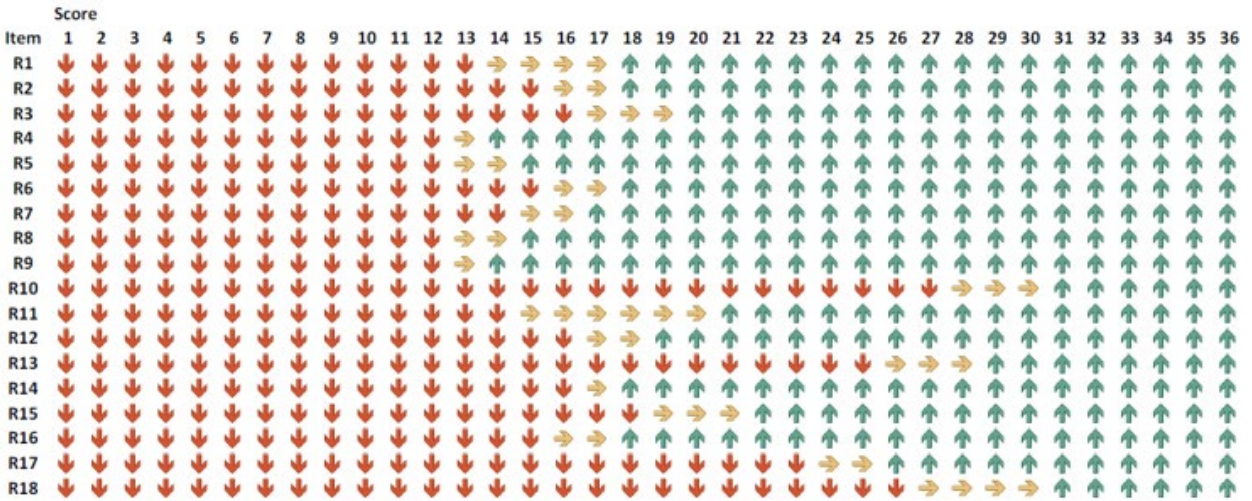
Day 2 (Wednesday)

Round 1 (ACT Modified Angoff Yes/No)

On Wednesday, each panel welcomed one or two additional panelists from the grade 8 standard setting. After a brief review/debrief, each subject panel participated in the Round 1 judgments, providing Yes/No judgments for borderline Approaching Proficient, Proficient, and Above Proficient for each item. After Round 1, feedback was provided in several forms. Panelists were provided with printouts of their individual judgments for each item (see excerpt in Table 34) and conditional p-values of the test (see excerpt in Figure 7). The facilitators presented aggregate information about panelists' individual raw cut score judgments (number of "Yes" responses for each performance level; see example from English in Table 35), as well as dot plots (see example in Figure 9) and summary statistics for the Round 1 cut scores on the 1–36 score scale (mean, median, minimum, maximum, Q1, and Q3; see Table 36). Panelists were also presented a table containing item-level percent agreement statistics so they could identify and discuss items with low percentages of agreement (see Table 37 for an excerpt from English). The full set of results tables and figures can be found in the Appendix.

Table 34. Excerpt of Printout of Individual Judgments for ACT Round 1

Item	Approaching	Proficient	Above
1	Yes	Yes	Yes
2	No	No	Yes
3	No	Yes	Yes
4	No	No	Yes
5	Yes	Yes	Yes
6	Yes	Yes	Yes
7	No	Yes	Yes
8	Yes	Yes	Yes
9	No	No	No
10	Yes	Yes	Yes



Display each icon according to these rules:

when value is
 when < 0.55 and
 when < 0.45

Figure 6. Excerpt of Conditional P-Values for ACT

Table 35. Round 1 Individual Raw Scores: ACT English

Subject	Panelist	Approaching	Proficient	Above
English	E01	27	52	66
English	E02	24	57	74
English	E03	56	74	75
English	E04	7	31	74
English	E05	8	39	75
English	E08	22	37	73
English	e09	34	60	75
English	E11	6	49	75
English	E12	59	73	75

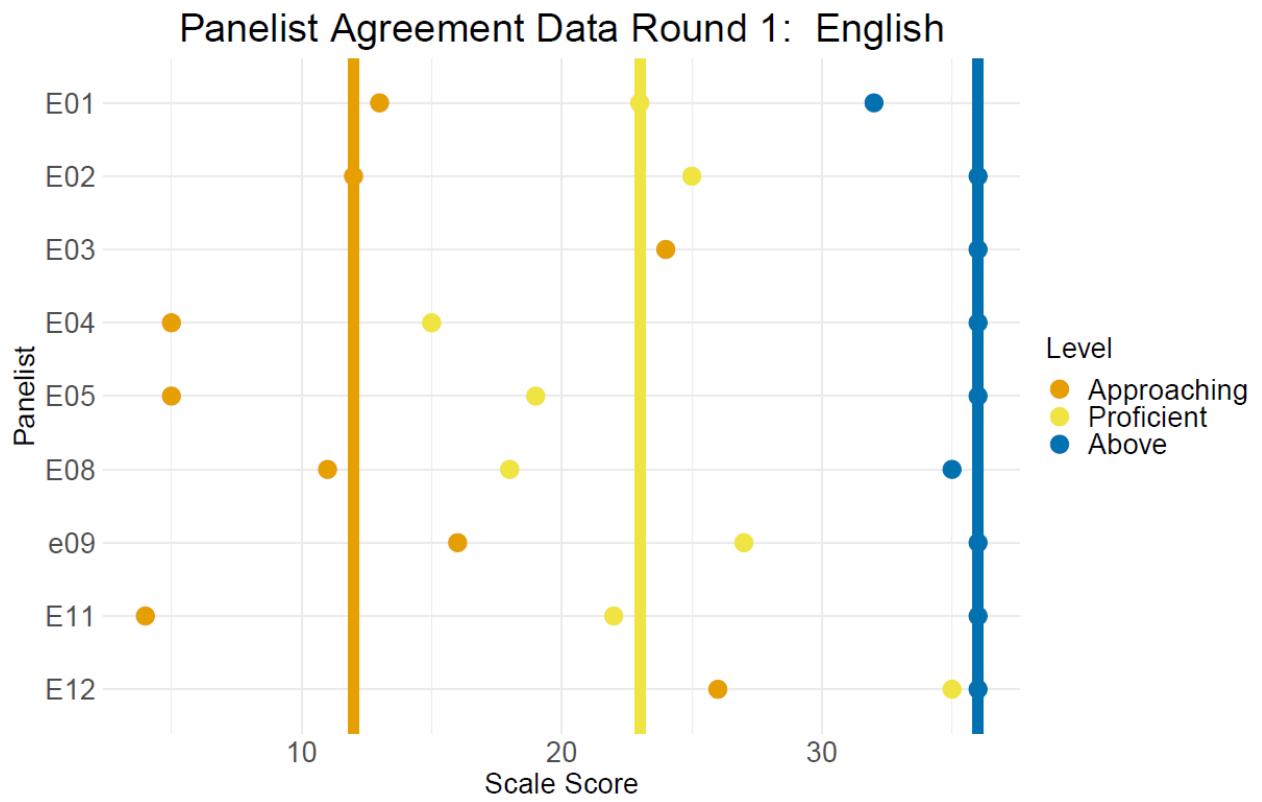


Figure 7. Round 1 Dot Plot for ACT English

Table 36. Score Statistics: Scale Scores

	Level	Number of Panelists	Min	Q1	Median	Mean	Q3	Max
Math	Above	14	30	31	32	33	35	36
	Proficient	14	17	21	24	24	26	30
	Approaching	14	6	13	13	14	16	20
Science	Above	10	27	32	36	33	36	36
	Proficient	10	19	21	25	25	28	36
	Approaching	10	8	10	14	14	17	24
English	Above	9	32	36	36	35	36	36
	Proficient	9	15	19	23	24	27	36
	Approaching	9	4	5	12	13	16	26
Reading	Above	11	26	32	35	33	36	36
	Proficient	11	18	22	24	24	26	32
	Approaching	11	7	12	14	14	16	18
Writing	Above	10	23	24	25	26	26	31
	Proficient	10	18	20	20	20	22	23
	Approaching	10	12	14	15	15	17	18

Table 37. Excerpt of Round 1 Percent Agreement of Item Judgments: English

Round 1 % Agreement of Item Judgments: English										
Subject	Item	Percent Yes Approaching	Percent Yes Proficient	Percent Yes Above	Percent Approaching: Yes, Yes, Yes	Percent Proficient: No, Yes, Yes	Percent Above: No, No, Yes	Percent No: No, No, No	Max Agreement	Performance Level
English	1	0.67	1.00	1.00	0.67	0.33	0.00	0.00	0.67	L2
English	2	0.11	0.67	1.00	0.11	0.56	0.33	0.00	0.56	L3
English	3	0.33	0.78	1.00	0.33	0.44	0.22	0.00	0.44	L3
English	4	0.22	0.67	1.00	0.22	0.44	0.33	0.00	0.44	L3
English	5	0.33	1.00	1.00	0.33	0.67	0.00	0.00	0.67	L3
English	6	0.56	0.89	1.00	0.56	0.33	0.11	0.00	0.56	L2
English	7	0.22	0.67	1.00	0.22	0.44	0.33	0.00	0.44	L3

Round 2 (ACT Modified Angoff Yes/No)

After an initial discussion of the Round 1 feedback, panelists reconvened in the general session for an introduction to a subset of the data that would be the main focus during Rounds 3 and 4. Panelists were shown the impact data for Illinois spring 2025 ACT-tested students (percentage of students scoring at or above each ACT score point in each subject area; see Figure 10 for example from English), and were provided an overview of the data underlying the ACT College Readiness Benchmarks, as well as a summary of how ACT scores are being used for admissions, course placement, and scholarships in 2-year and 4-

year colleges and universities in Illinois. Panelists were provided a handout containing the probabilities of success and 2025 Illinois impact data, along with instructions for how to find the Round 1 cut scores on the handout (see Table 38 for an example from English; the Round 1 handouts for each subject area can be found in the Appendix). Panelists then returned to their breakout rooms for additional discussion before providing their Round 2 judgments. Round 2 also used the Modified Angoff Yes/No approach. After a discussion of the Round 2 results, including the same types of data that were shared after Round 1, panelists completed a second evaluation and were dismissed for the day.

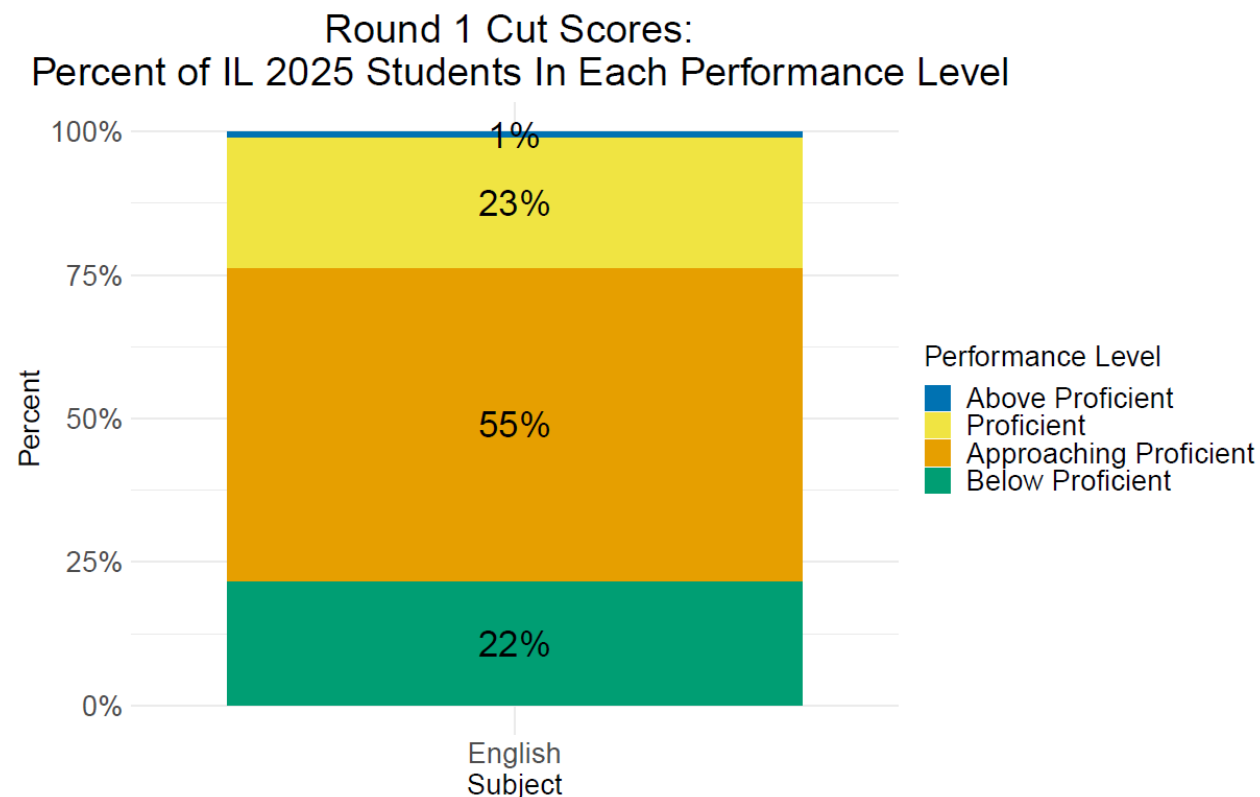


Figure 8. Round 1 Impact Data, English

Table 38. Round 1 Cut Scores with Probabilities of Success and Impact: English

Subject	Level	ACT Score	Probability			Percentage At or Above
			A	B or higher	C or higher	IL Jr 2025
English	Above	36	0.78	0.93	0.96	1
English	Proficient	23	0.28	0.68	0.87	24
English	Approaching	12	0.06	0.33	0.69	78

Day 3 (Thursday)

Round 3 (ACT Briefing Book)

On Thursday the ACT panels reconvened and transitioned to the empirical/briefing book approach. A subset of panelists from the English, reading, and writing panels were combined into a single ELA panel; the math panel remained the same as the previous day, and the science panel welcomed an additional panelist to represent workforce development.

After a brief review and discussion of the Round 2 results, panelists returned to the general session for an introduction of the Briefing Book data. Panelists were provided with briefing books containing information specific to their subject area and the lead facilitator walked panelists through each page of the data book to help them understand the data and any limitations or caveats to the data. The information in the briefing books was tied to the Round 2 results where applicable. For example, panelists were shown how they could find the probabilities of earning an A, B or higher, or C or higher grade associated with the Round 2 results, the percent of students who enrolled in college associated with the Round 2 results, and the PreACT Secure scores in grades 9 and 10 associated with the Round 2 results based on an equipercentile linkage. After reviewing the briefing books, panelists returned to the breakout rooms for additional discussion.

For Round 3, panelists received training about how to set cut scores using the Briefing Book approach. Panelists were instructed to think about minimally Proficient students in their subject area, and think about their likelihood of earning an A, B or higher, or C or higher grade in a first-year credit-bearing college course in that subject area. Using the page in their briefing books containing the ACT scores, probabilities of success, and impact data, panelists were instructed to select the ACT score that best reflects what they considered minimally Proficient in terms of the probability of earning an A, B or higher, or C or higher grade. This task was repeated for the Approaching Proficient and Above Proficient cut scores, resulting in three judgments from each panelist. Panelists provided their individual judgments before breaking for lunch.

Round 4 (ACT Briefing Book)

After lunch, panelists engaged in discussion in the breakout rooms before reconvening in the general session. In the breakout rooms, panelists were asked to consider how their Round 2 cut scores compared to their Round 3 cut scores, and to consider how the additional information in the briefing books impacted their Round 3 judgments, and which information was more important or compelling in their cut score determinations. In the general session, panelists were given the opportunity to share discussion topics across the three subject areas. Panelists were shown a comparison of the results of Round 2

compared to Round 3 across subject areas and provided an opportunity to discuss any discrepancies across rounds or across subject areas. Panelists were asked to consider all of the information they had reviewed during the standard setting, including the PLDs, the test content, and the data contained in the briefing books when making their final cut score recommendations.

Panelists returned to the breakout rooms to enter their final cut score recommendations using the same approach as Round 3, and to complete a final process evaluation. After a short break, panelists returned to the general session for a presentation of the final results as well as closing comments by ISBE. Panelists were thanked for their participation in the standard setting and then were dismissed.

The final ACT cut score recommendations were also linked to PreACT 9 Secure and PreACT Secure scores, using an equipercentile linking approach. For each recommended cut score on the ACT, the percent of students who took the ACT in Illinois as part of spring statewide testing scoring at or above that score in the Illinois spring 2025 data was calculated, then that percentile was used to find the PreACT 9 Secure score and PreACT Secure score with the same percent of students who scored at or above that score in the grade 9 and grade 10 Illinois spring 2025 statewide tested data. The resulting cut scores were presented during the vertical articulation and were approved or adjusted by the panelists during the vertical articulation process.

Vertical Articulation (Day 5)

Following the standard setting, a subset of panelists participated in a vertical articulation meeting for each subject to review the final cut score recommendations along with student impact data to consider whether and to what extent adjustments to the recommended cut scores were warranted based on content and policy considerations. The participants for the vertical articulation had been selected during the recruitment process. All participants knew which committees they would be serving on throughout the week and whether they were pre-selected to stay for the final day for Vertical Articulation.

After a general session provided an overview of the vertical articulation purpose and process, each subject-level committee reviewed PLDs across the grades to determine differences in expectations and predicted the impact distribution across grades based on these expectations. The facilitator for the articulation meeting led the panelists through a discussion related to the expectations across grades. The committee reviewed the articulation of the recommended cut scores based on the Spring 2025 impact data, where final adjustments could have been recommended to ensure coherent expectations across grade levels.

Panelists determined what, if any, adjustments were needed and supported. If the panelists discovered a misalignment in the impact data or recommended ranges, they had an opportunity to discuss and possibly adjust the committee cut score recommendations. Panelists were able to observe the effect of changes to the cut scores on the impact data using an interactive impact data spreadsheet. The facilitator imputed recommended cut score changes from the articulation committee and displayed the resulting impact data. Based on their rationale and adjustments, the committee then determined whether to recommend changes or to retain the final cut score recommendations from each committee. The Mentimeter app was used to collect panelist agreement with each of the proposed changes until consensus was achieved.

Any recommended adjustments to the cut scores were kept within the range of cut score recommendations from the final standard setting judgments, and all vertical articulation panelists were encouraged to support the final recommended cut scores. In general, the goal was for discussion to continue until consensus was reached. Once the content area panels had completed their discussions, all content areas were brought together for a final discussion of the vertical articulation results across content areas to ensure coherence across content areas. If any content area results seemed inconsistent with the other content area results, additional discussion ensued in a similar fashion to the content area discussions. A final vote using the Mentimeter app indicated consensus agreement with the final recommended cut scores.

Panelists completed an evaluation of the articulation process and meeting on the standard setting website.

Results

Results for Grades 3–8

IAR ELA

During the standard setting process for ELA, panelists engaged in three rounds of ratings using the Extended Modified Angoff Yes/No Method. Cuts were presented on the scaled metric. The scale for ELA is from 650 to 850 with historical cuts at 700, 750, and near 800 for the performance levels that are would roughly map onto the new category names of Approaching Proficient, Proficient, and Above Proficient, respectively. Grades 4, 6, and 8 met first, on Monday and Tuesday. In Round 1, panelists tended to set cut scores that were relatively high compared to historical benchmarks, particularly for Above Proficient and Proficient. This is evidenced by the median cut scores for Above Proficient (841 for grade 4, 814 for grade 6, and 850 for grade 8), and for Proficient (779, 779, and 786). The Approaching Proficient cuts (722, 739, and 721) were more similar to the original cut of 700 for that level. (It should be noted that the old scale for the IAR had 5 performance levels and three performance levels fell below the Proficient cut. In essence, this allowed for both the original cut for Level 2 at 700 and the original cut for Level 3 at 725 to represent two levels within the scale area that is most similar to the current "Approaching Proficient" cut.)

These initial judgments were likely the result of overestimating the most common rubric score for students just entering each performance level, a common outcome when participants are first introduced to the standard setting process and must rely solely on their understanding of the descriptors and item performance. The range of scores was fairly wide across all levels in Round 1, reflecting variability in interpretation across panelists.

In Round 2, panelists were presented with student profile data, which gave them a clearer picture of how real students performed at each score point. This additional information helped ground their judgments and resulted in more moderate recommendations. Medians for Above Proficient decreased to 800 (grade 4), 772 (grade 6), and 812 (grade 8). Likewise, Proficient medians dropped to 754, 736, and 769, and Approaching Proficient medians declined to 697, 692, and 710. In addition to this downward shift, the spread of scores narrowed, particularly in the Approaching Proficient range, indicating growing alignment among panelists.

By Round 3, the panelists had greater confidence in their understanding of the performance levels and expectations, and their ratings showed increased consistency. Median cut scores for Above Proficient were 788 (grade 4), 795 (grade 6), and 783 (grade 8); for

Proficient, 752, 739, and 758; and for Approaching Proficient, 675, 702, and 704. These scores remained relatively stable from Round 2, with only minor adjustments. The interquartile ranges narrowed further, suggesting that consensus had been reached and that the group was comfortable with the final recommendations.

Later in the week, on Wednesday and Thursday, Grades 3, 5, and 7 were reviewed. These panels benefited from both improved process familiarity and insights from prior groups. As a result, Round 1 medians for these grades were more moderate than those seen earlier in the week. For Above Proficient, the medians were 795 (grade 3), 793 (grade 5), and 782 (grade 7); for Proficient, 749, 749, and 747; and for Approaching Proficient, 685, 692, and 716. These scores were already closer to historical cut points, and the spread of values was narrower, indicating greater initial alignment among panelists.

In Round 2, slight downward adjustments were made as panelists reviewed student performance profiles, with Above Proficient medians decreasing to 790 (grade 3), 793 (grade 5), and 778 (grade 7); Proficient medians shifting to 746, 742, and 747; and Approaching Proficient medians settling at 682, 669, and 716. By Round 3, the medians remained largely stable: 788, 795, and 783 for Above Proficient; 749, 744, and 746 for Proficient; and 685, 700, and 716 for Approaching Proficient. The increasingly narrow score distributions indicated strong convergence and agreement among panelists. Overall, the process allowed panelists to refine their expectations and arrive at valid, empirically informed cut scores aligned to performance level descriptors and real student achievement. See Table 39 for the full set of scaled score cuts for Rounds 1-3 for ELA. Figures 11-16 show the final panelist scaled score ratings after Round 3 for grades 3-8. Figures 17-18 show the Impact data after Round 2 and Round 3.

Table 39. Scaled Score Results by Round for ELA

Round 1	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
ELA3	766	788	795	799	823	700	736	749	754	769	650	650	685	691	705
ELA4	778	825	841	850	850	728	766	779	797	816	650	715	722	728	743
ELA5	782	791	793	808	824	730	740	749	765	772	650	650	692	707	713
ELA6	784	811	814	817	838	754	764	779	790	794	714	723	739	745	747
ELA7	755	763	782	787	813	728	741	747	760	775	693	701	716	723	737
ELA8	818	837	850	850	850	751	775	786	795	837	689	713	721	731	764
Round 2	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
ELA3	762	783	790	797	827	730	744	746	749	762	661	666	682	685	696
ELA4	780	794	800	810	836	737	748	754	766	780	650	678	697	712	752
ELA5	784	786	793	797	806	733	733	742	751	760	650	650	669	682	696
ELA6	757	768	772	778	784	708	732	736	747	763	650	684	692	708	711
ELA7	763	773	778	787	802	737	741	747	752	775	688	693	716	721	733
ELA8	771	798	812	820	850	735	758	769	781	833	650	701	710	732	747
Round 3	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
ELA3	762	783	788	797	827	739	746	749	752	752	650	675	685	688	716
ELA4	780	789	797	808	822	739	746	752	760	780	650	650	675	699	754
ELA5	777	789	795	797	806	733	737	744	749	760	650	687	700	710	713
ELA6	768	771	773	777	780	721	734	739	743	757	678	698	702	711	721
ELA7	768	780	783	787	805	737	741	746	752	776	688	697	716	721	733
ELA8	744	795	802	807	829	710	753	758	764	825	650	693	704	710	742

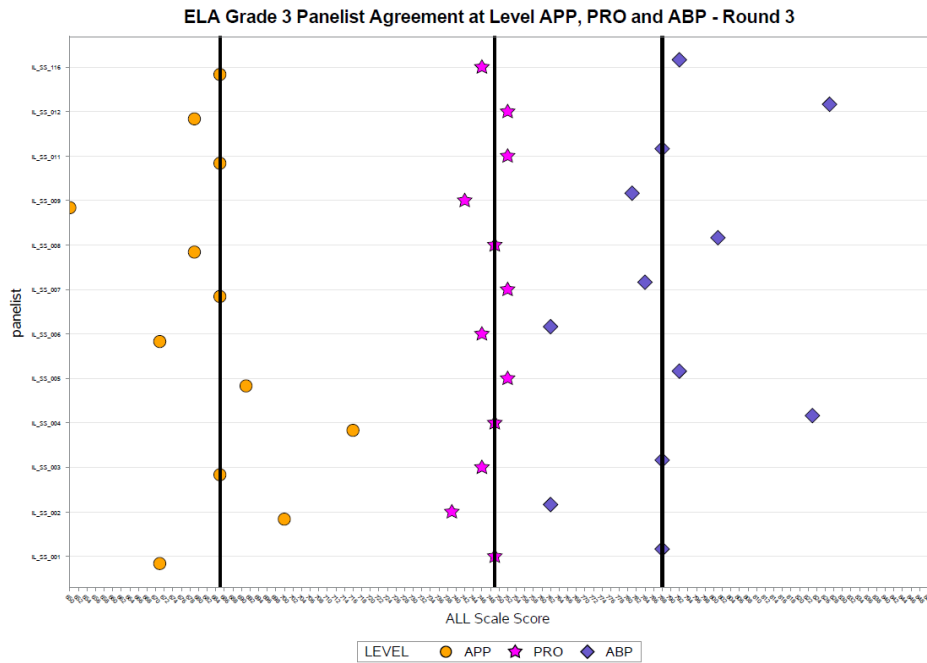


Figure 9. ELA Grade 3 Panelist Agreement

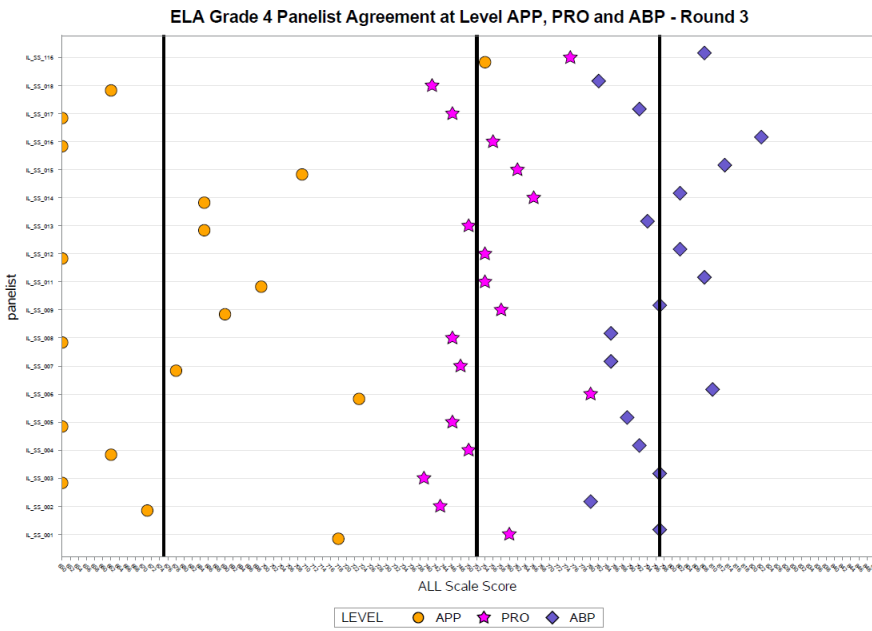


Figure 10. ELA Grade 4 Panelist Agreement

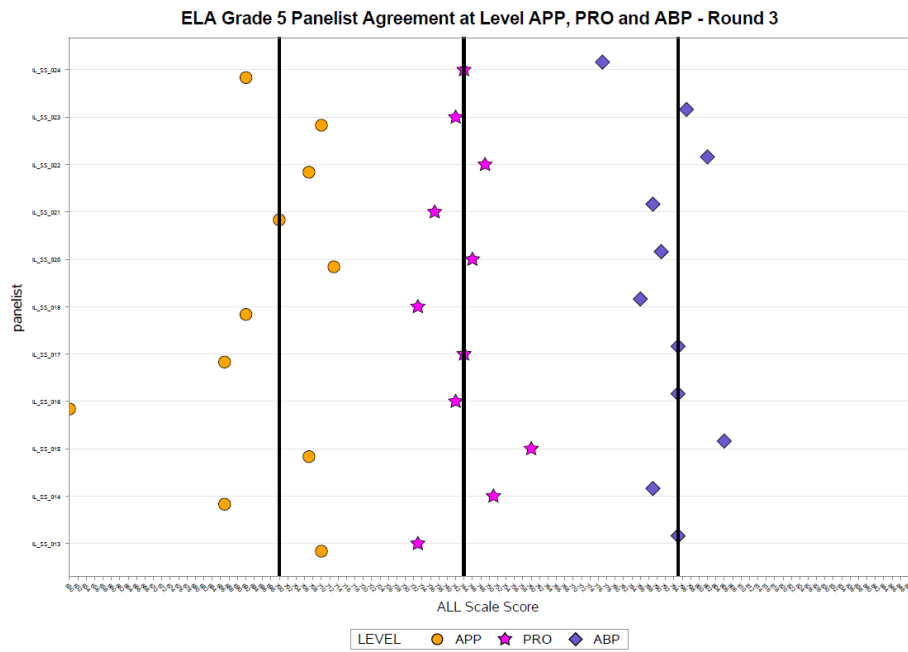


Figure 11. ELA Grade 5 Panelist Agreement.

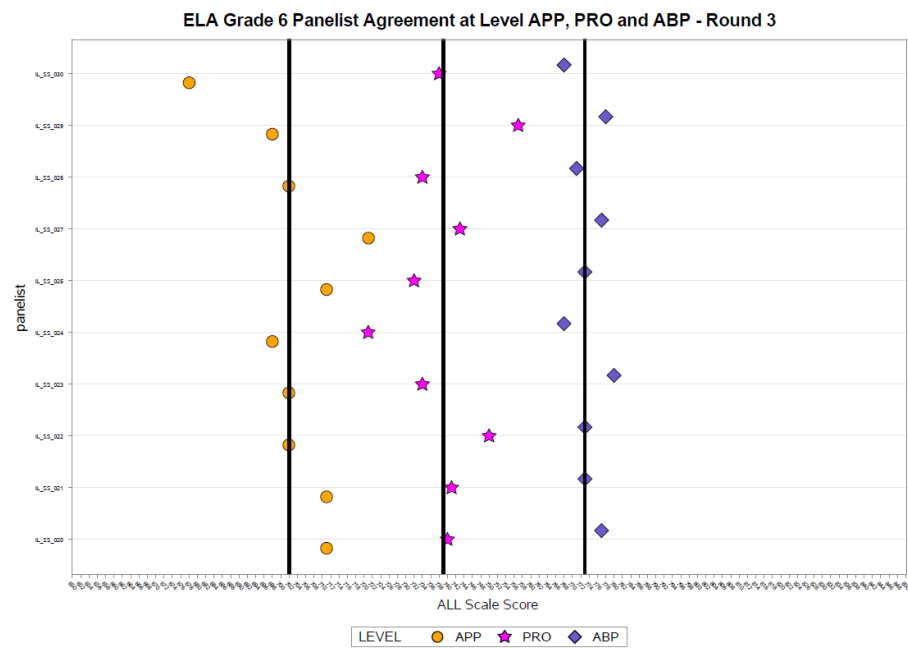


Figure 12. ELA Grade Panelist Agreement.

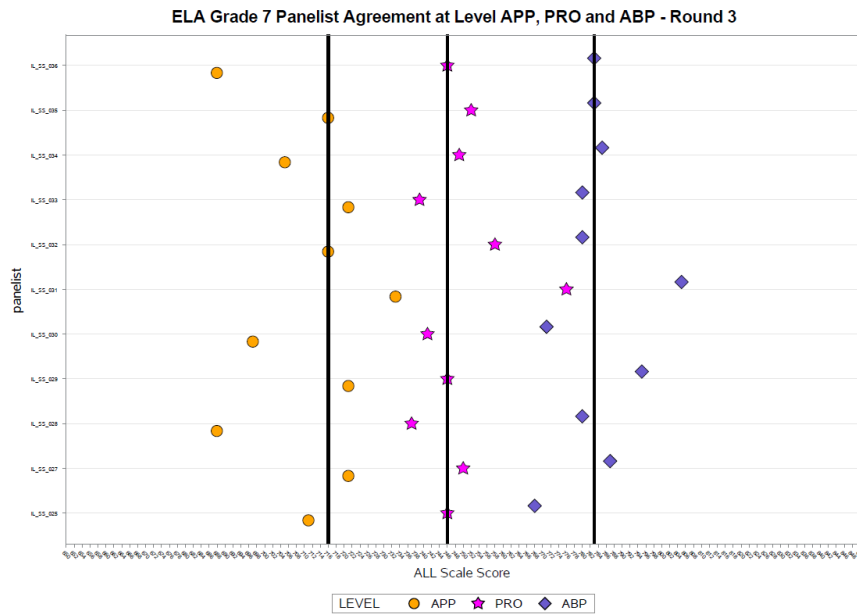


Figure 13. ELA Grade 7 Panelist Agreement.

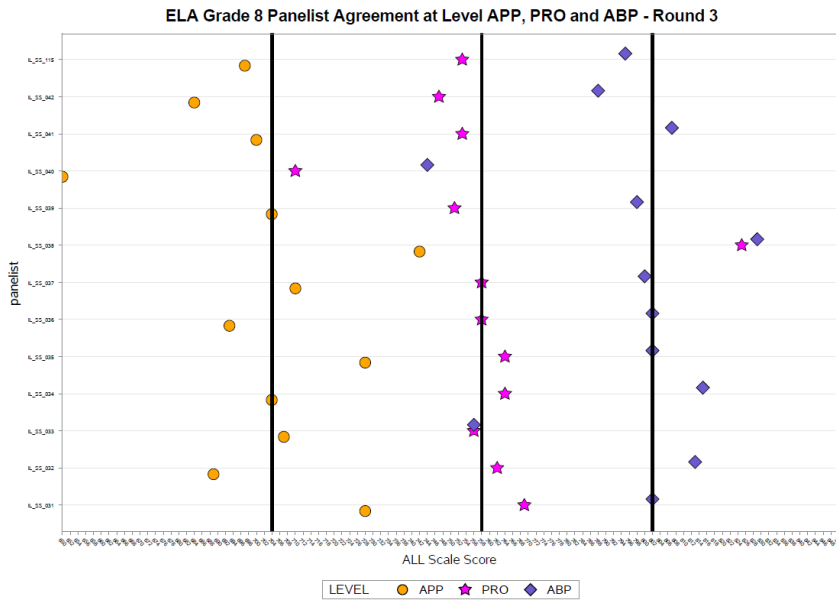


Figure 14. ELA Grade 8 Panelist Agreement.

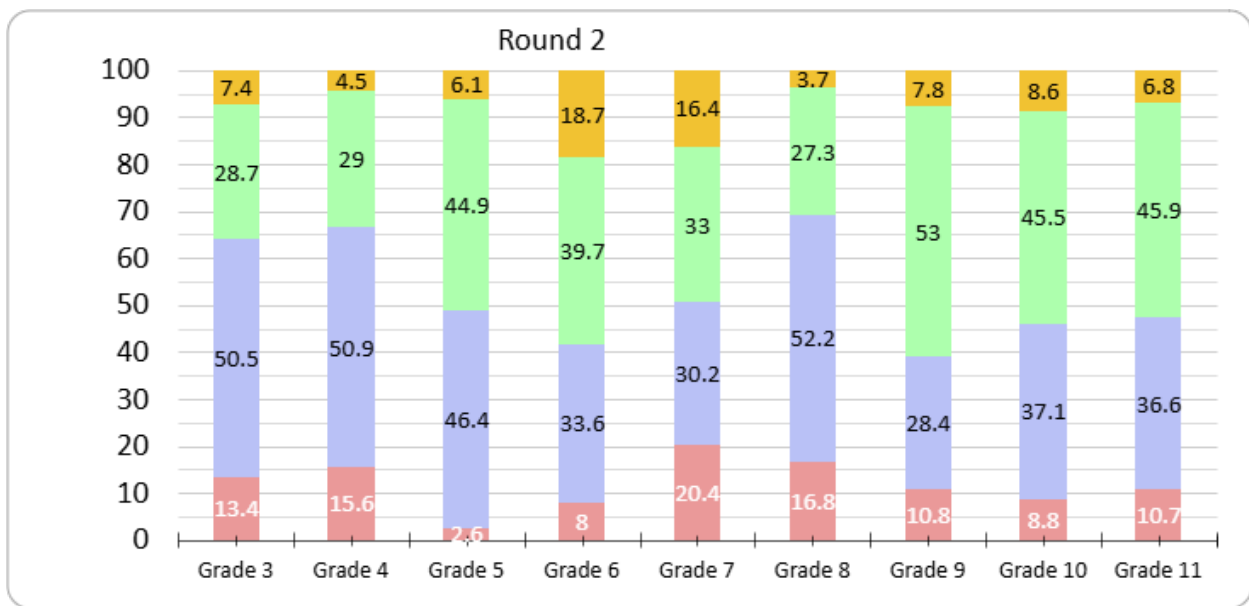


Figure 15. ELA Impact Data: Round 2

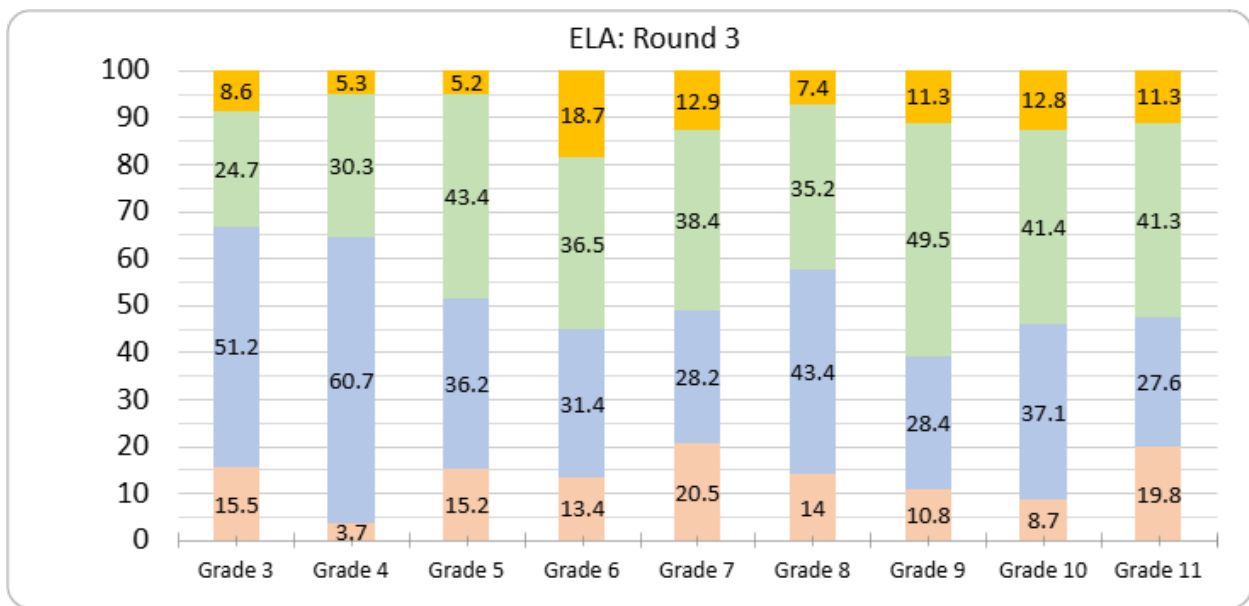


Figure 16. ELA Impact Data: Round 3

IAR Math

During the standard setting process for Mathematics, panelists engaged in three rounds of ratings using the Extended Modified Angoff Yes/No Method. As with ELA, Grades 4, 6, and 8 met first on Monday and Tuesday, followed by Grades 3, 5, and 7 on Wednesday and Thursday. In Round 1, panelists for Grades 4, 6, and 8 tended to overestimate student performance at the lower end of each performance level, resulting in relatively high median cut scores. For the Above Proficient level, the median scores were 801 for Grade 4, 847 for Grade 6, and 850 for Grade 8. Proficient medians were also elevated at 756 (Grade 4), 781 (Grade 6), and 795 (Grade 8). The Approaching Proficient medians were 710 for Grade 4, 741 for Grade 6, and 721 for Grade 8. These scores, particularly in the Proficient and Above Proficient ranges, were well above historical expectations and reflected a rigorous interpretation of the performance descriptors. The wide spread in Round 1 scores reflected variability in understanding of the performance level descriptors across the panel.

In Round 2, panelists reviewed student profile data to better understand actual student performance. With this information, they recalibrated their judgments. As a result, median scores dropped across all performance levels. Above Proficient medians decreased to 782 for Grade 4, 785 for Grade 6, and 784 for Grade 8. Proficient medians fell to 738, 752, and 725, respectively, and Approaching Proficient medians decreased to 669, 702, and 650. These shifts indicate that panelists shifted their understanding of student performance and adjusted their ratings accordingly. Score spreads, especially in the Approaching and Proficient levels, became narrower, suggesting increased alignment and shared expectations among panelists. However, panelists realized that they overcorrected as their cutscore for Approaching Proficient fell to the lowest obtaining scaled score of 650.

By Round 3, panelist ratings had largely stabilized, with only minor refinements. Median cut scores for Above Proficient were 784 (Grade 4), 773 (Grade 6), and 791 (Grade 8). Proficient medians came in at 740, 742, and 746, and Approaching Proficient medians finished at 708, 702, and 684. The narrower interquartile ranges demonstrated strong convergence in panelist judgments. Panelists grew increasingly confident in their understanding of performance levels, supported by both the process and empirical data.

Later in the week, Grades 3, 5, and 7 met for their standard setting activities. These panels benefited from observing the process earlier in the week and approached Round 1 with greater clarity. Their initial medians were more moderate than those of the earlier groups. In Round 1, Above Proficient medians were 805 (Grade 3), 795 (Grade 5), and 780 (Grade 7). Proficient medians were 761, 760, and 747, and Approaching Proficient medians were 715, 721, and 712, respectively. These medians were closer to historical cut points and demonstrated stronger initial consensus than Grades 4, 6, and 8. The interquartile ranges were also narrower than the first set of grades.

In Round 2, panelists refined their ratings after reviewing student profile data. Above Proficient medians shifted downward to 774 (Grade 3), 783 (Grade 5), and 779 (Grade 7). Proficient medians decreased slightly to 732, 744, and 746, while Approaching Proficient medians adjusted to 697, 704, and 712. These shifts reflect careful consideration of how the descriptors aligned with student performance data and further alignment among panelists.

By Round 3, panelist ratings showed only minor changes, indicating confidence and agreement. Above Proficient medians were 771 (Grade 3), 782 (Grade 5), and 780 (Grade 7). Proficient medians held close to the previous round at 732, 748, and 745, while Approaching Proficient medians were 700, 709, and 712, respectively. The tight clustering of scores and consistency across Rounds 2 and 3 reinforced the reliability of the final cut score recommendations.

Overall, the Mathematics standard setting process showed a clear pattern of initial overestimation in early grades, followed by refinement and convergence across rounds. Grades 3, 5, and 7 began with more moderate medians due to improved process familiarity. Across all grades, Round 3 reflected strong consensus as evidenced by the tight interquartile range and plots of final panelist cuts. See Table 40 for the full set of scaled score cuts for Rounds 1-3 for Math. Figures 19-24 show the final panelist scaled score ratings after Round 3 for grades 3-8. Figures 25-26 show the Impact data after Round 2 and Round 3.

Table 40. Scaled Score Results by Round for Math

R1	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
MATH3	789	800	805	814	850	746	755	761	770	797	677	702	715	726	738
MATH4	773	785	801	813	850	736	751	756	765	776	658	689	710	718	731
MATH5	775	783	795	813	826	736	749	760	768	778	688	699	721	731	744
MATH6	768	816	847	850	850	745	779	781	788	802	682	719	741	749	760
MATH7	766	772	780	793	800	735	745	747	756	774	687	704	712	720	748
MATH8	822	827	850	850	850	741	778	795	813	850	688	688	721	741	805
R2	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
MATH3	765	771	774	774	777	723	728	732	735	740	693	695	697	700	702
MATH4	765	779	782	788	792	726	733	738	745	749	650	650	669	693	712
MATH5	781	781	783	783	804	741	744	744	751	753	699	704	704	713	716
MATH6	774	779	785	792	850	739	743	752	762	782	672	702	702	711	747
MATH7	766	772	779	789	795	735	744	746	752	759	695	710	712	719	733
MATH8	748	767	784	802	839	672	680	725	744	772	650	650	650	680	688
R3	Above Proficient					Proficient					Approaching Proficient				
Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
MATH3	765	770	771	771	777	726	729	732	735	738	693	697	700	702	706
MATH4	776	782	784	788	792	731	736	740	747	756	701	705	708	712	723
MATH5	778	781	782	786	796	741	744	748	753	753	699	704	709	713	716
MATH6	755	770	773	776	828	707	739	742	745	759	650	682	702	702	722
MATH7	766	773	780	789	795	735	743	745	753	759	695	708	712	717	733
MATH8	767	786	791	813	817	694	733	746	748	764	662	672	684	706	721

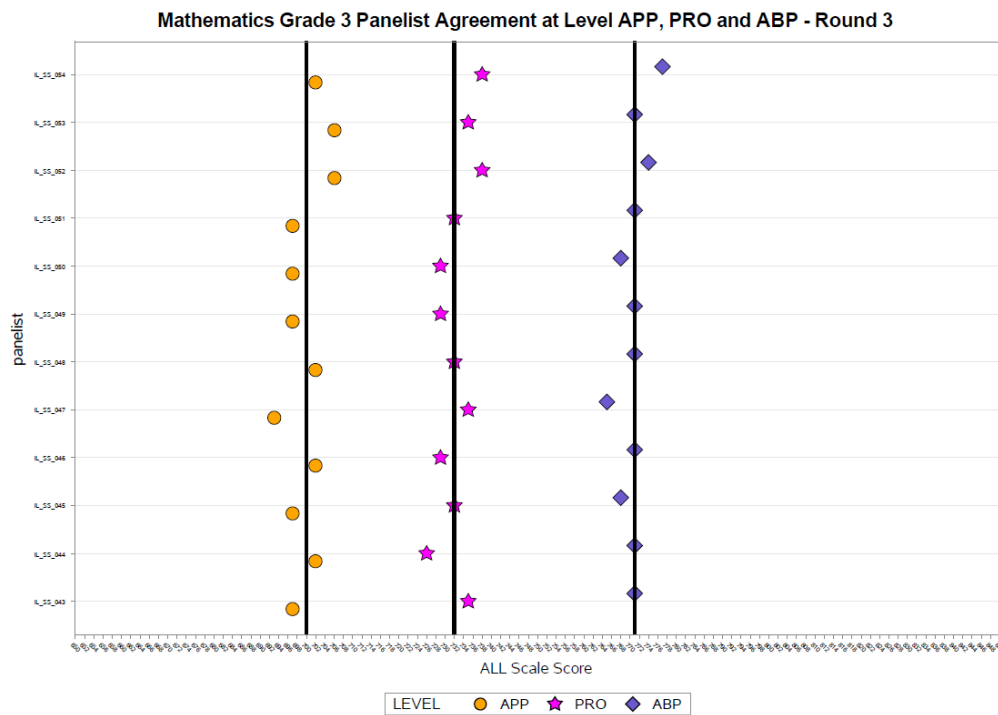


Figure 17. Mathematics Grade 3 Panelist Agreement

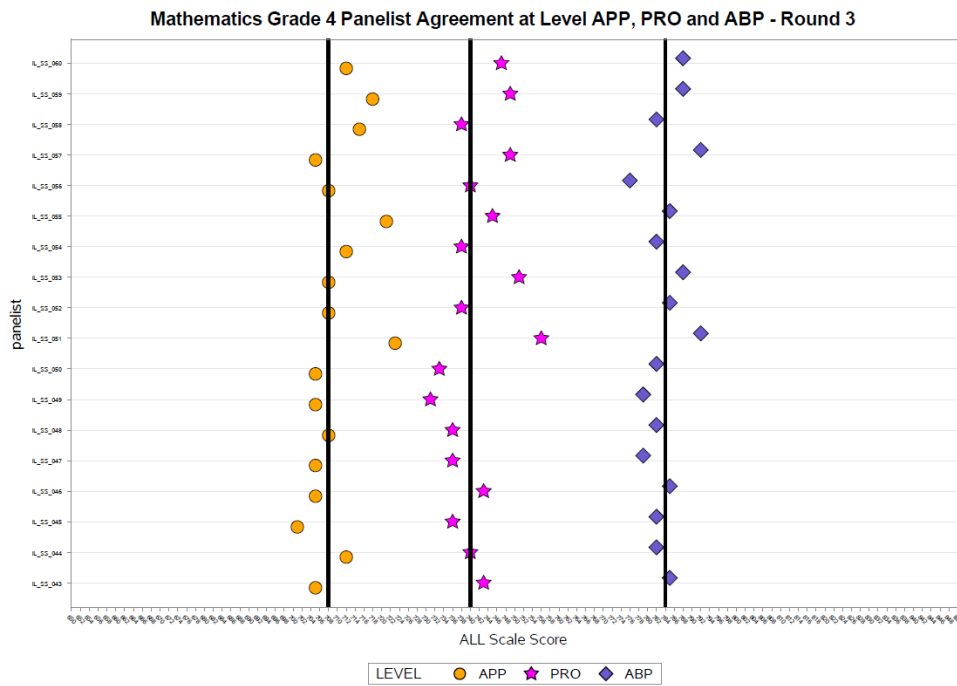


Figure 18. Mathematics Grade 4 Panelist Agreement

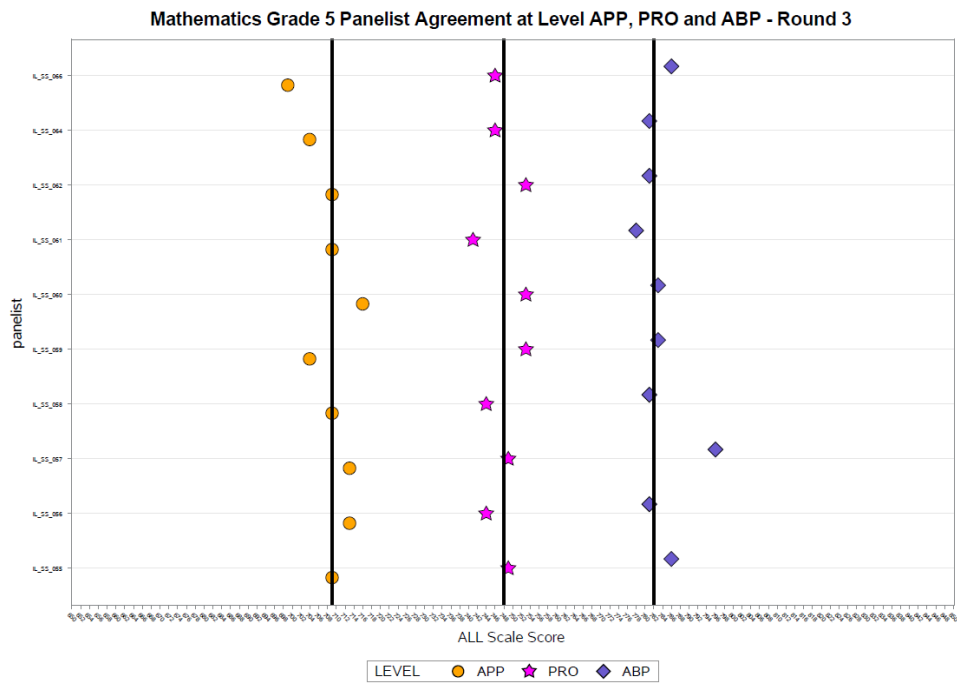


Figure 19. Mathematics Grade 5 Panelist Agreement

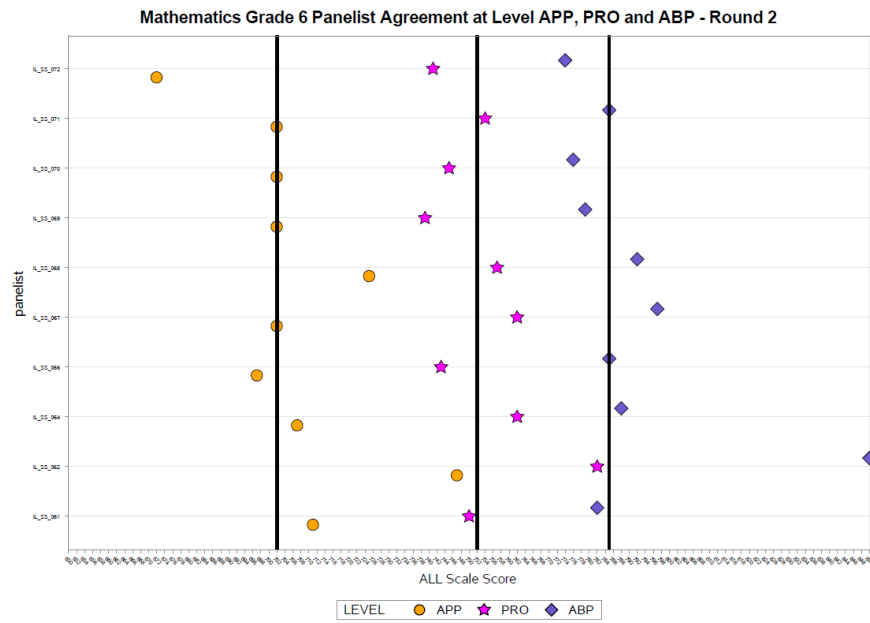


Figure 20. Mathematics Grade 6 Panelist Agreement

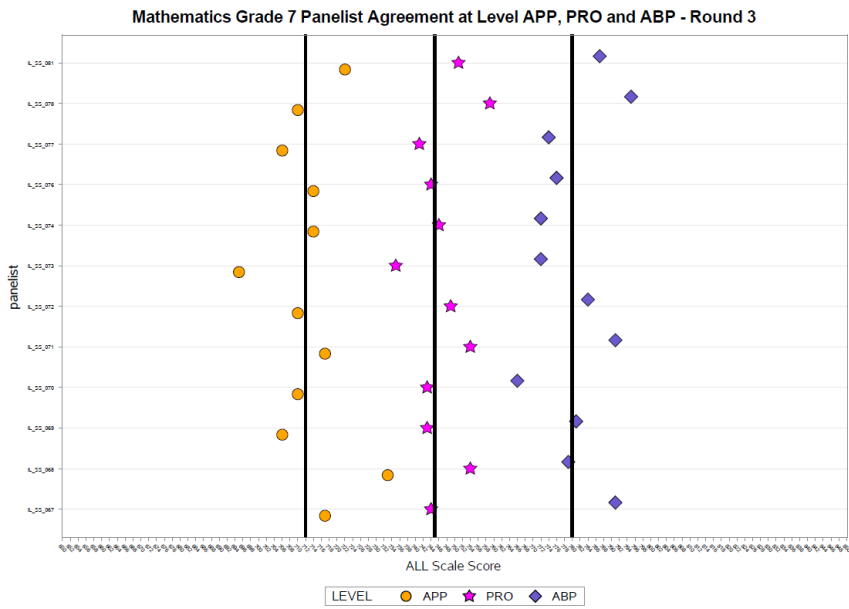


Figure 21. Mathematics Grade 7 Panelist Agreement

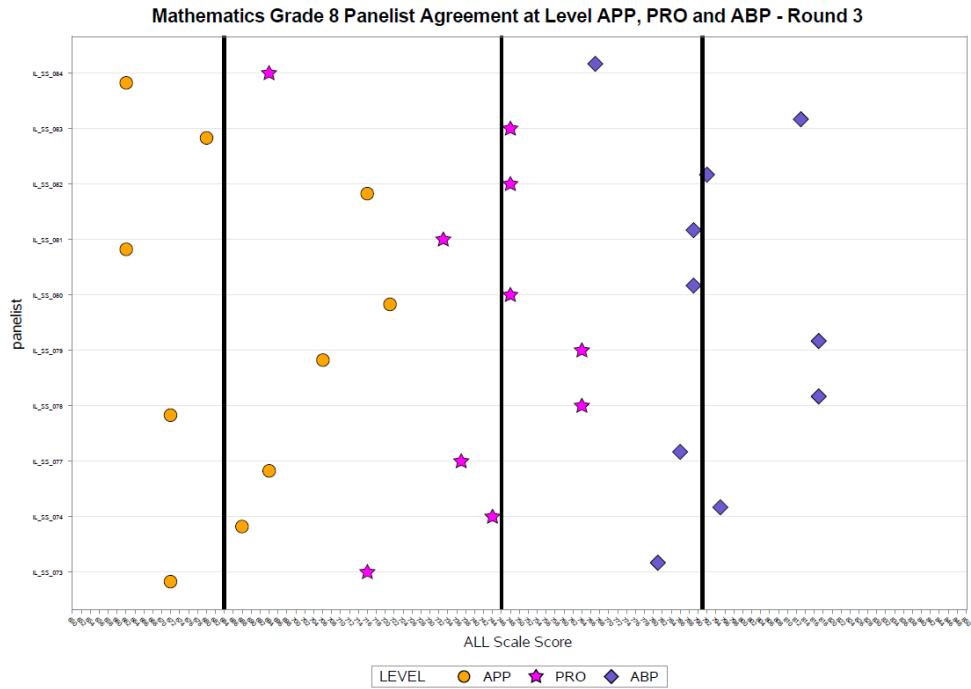


Figure 22. Mathematics Grade 8 Panelist Agreement

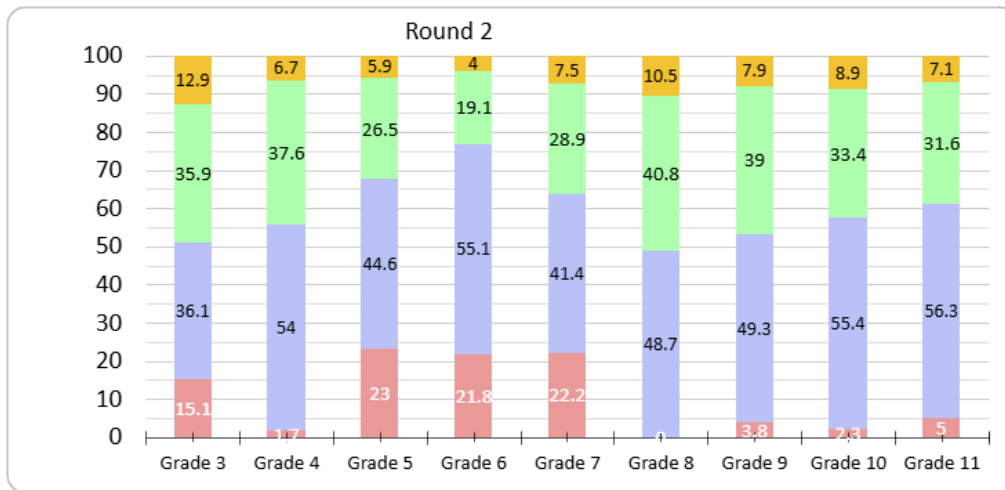


Figure 23. Math Impact Data: Round 2

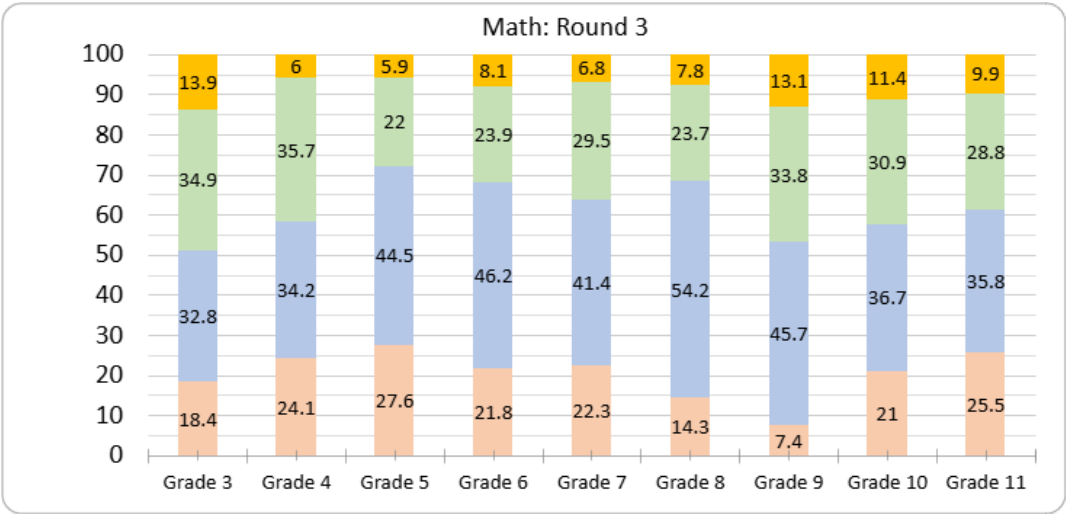


Figure 24. Math Impact Data: Round 3

ISA Science

The standard setting process for Science followed the same three-round structure as ELA and Math, using the Extended Modified Angoff Yes/No Method. Grade 8 (SCIE8) convened first, meeting on Monday and Tuesday. In Round 1, panelists tended to set very high cut scores, especially for the Above Proficient level, where the median was 900, the maximum possible scaled score. The reporting scale for science is from 700–900. This high cut in Round 1 indicates that many panelists were initially overestimating what students at the bottom of this performance level could reliably demonstrate. The median for Proficient was also high at 842, and the median for Approaching Proficient was 787. These medians exceeded typical expectations and reflected a rigorous interpretation of performance levels, prior to seeing any student data.

In Round 2, after reviewing student profile data, grade 8 science panelists revised their judgments to better align with actual student performance. The median for Above Proficient decreased slightly to 870, and the Proficient median dropped to 813, while the Approaching Proficient median declined to 758. These moderate adjustments suggest that panelists recalibrated their expectations based on evidence. By Round 3, medians remained relatively stable with only moderate adjustments: 876 for Above Proficient, 822 for Proficient, and 760 for Approaching Proficient. This stability across Rounds 2 and 3, along with narrower interquartile ranges, indicates strong convergence and confidence in the final recommended cut scores for Grade 8 Science.

Later in the week, the panel for Grade 5 Science (SCIE5) met on Wednesday and Thursday. Only three panelists from the Grade 8 panel participated in the Grade 5 process, given the difference in certification and classroom experience between middle school and elementary science educators. As a result, grade 5 science ratings evolved independently, though they followed a similar pattern of refinement across rounds. In Round 1, the median for Above Proficient was 861, for Proficient 811, and for Approaching Proficient 761. These medians, while still high, were slightly lower than those observed in Grade 8 and suggest a slightly more tempered starting point.

In Round 2, after reviewing student profile data, SCIE5 panelists maintained a relatively consistent view of Above Proficient performance, with the median increasing slightly to 865, while the Proficient median remained at 811, and the Approaching Proficient median decreased modestly to 757. By Round 3, panelists settled on 856 for Above Proficient, 804 for Proficient, and 766 for Approaching Proficient. These small refinements across Rounds 2 and 3 reflect thoughtful consensus among the Grade 5 panel and indicate confidence in the defensibility of their final recommendations.

Overall, the Science standard setting process resulted in high initial cut scores that were refined as panelists engaged more deeply with student data. While Grade 8 began with

exceptionally high medians, the process helped to moderate and clarify expectations. Grade 5 panelists started with slightly lower estimates and also refined their judgments through subsequent rounds. Despite limited overlap between panels, both groups reached consistent and well-informed conclusions by the end of Round 3. See Table 41 for scaled cuts for all rounds. Figures 27–28 show the final panelist scaled score ratings after Round 3 for grades 5 and 8. Figures 29–30 show the Impact data after Round 2 and Round 3.

Table 41. Scaled Score Results by Round for Science

R1	Above Proficient					Proficient					Approaching Proficient				
	Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3
SCIE5	829	846	861	897	897	781	799	811	838	856	713	755	761	772	780
SCIE8	845	900	900	900	900	806	820	842	862	898	751	770	787	808	824
R2	Above Proficient					Proficient					Approaching Proficient				
	Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3
SCIE5	843	846	865	885	897	794	799	811	817	840	741	746	757	763	765
SCIE8	818	851	870	900	900	764	796	813	845	889	700	747	758	774	813
R3	Above Proficient					Proficient					Approaching Proficient				
	Stat	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3
SCIE5	827	843	856	865	877	789	796	804	807	831	739	761	766	777	789
SCIE8	808	848	876	889	900	732	794	822	840	876	700	747	760	774	799

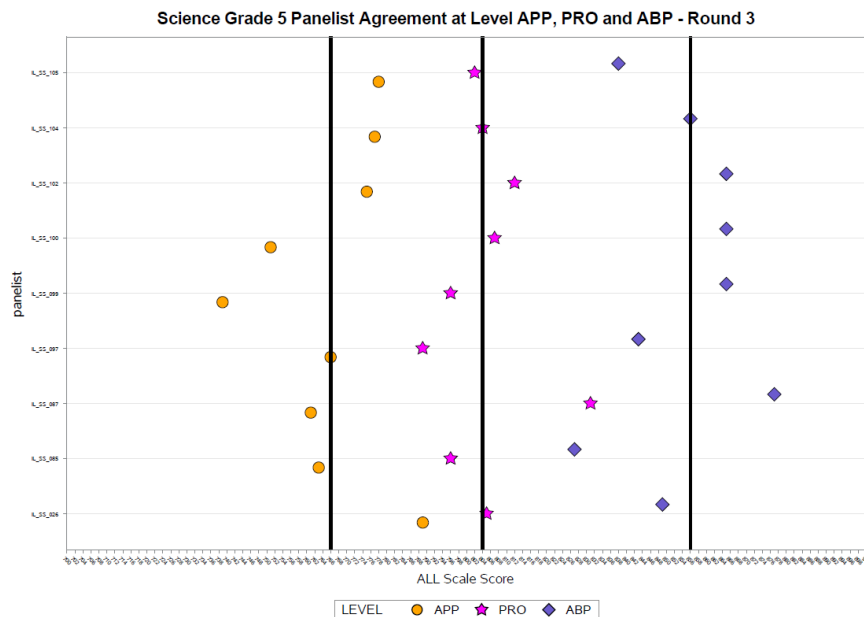


Figure 25. Science Grade 8 Panelist Agreement

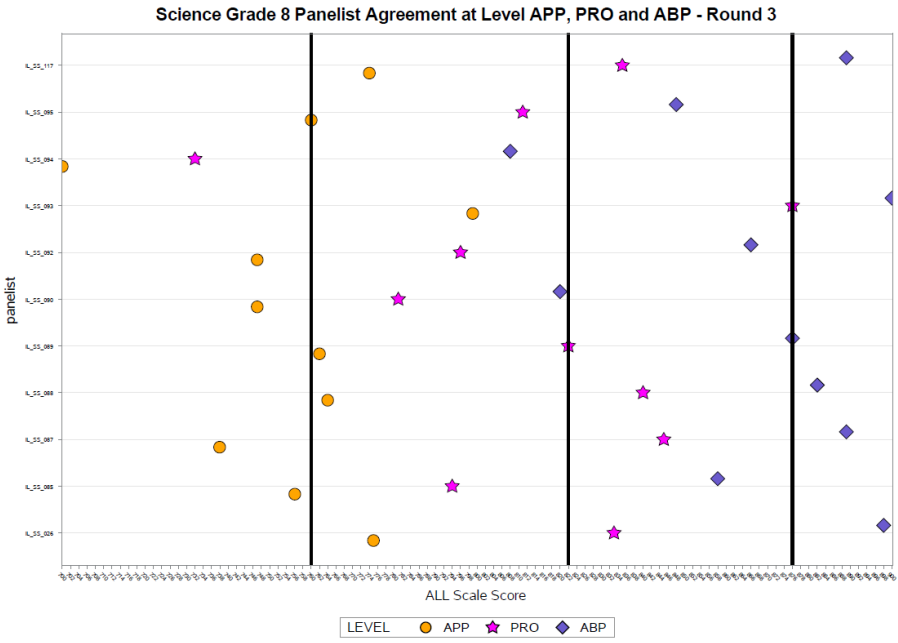


Figure 26. Science Grade 8 Panelist Agreement

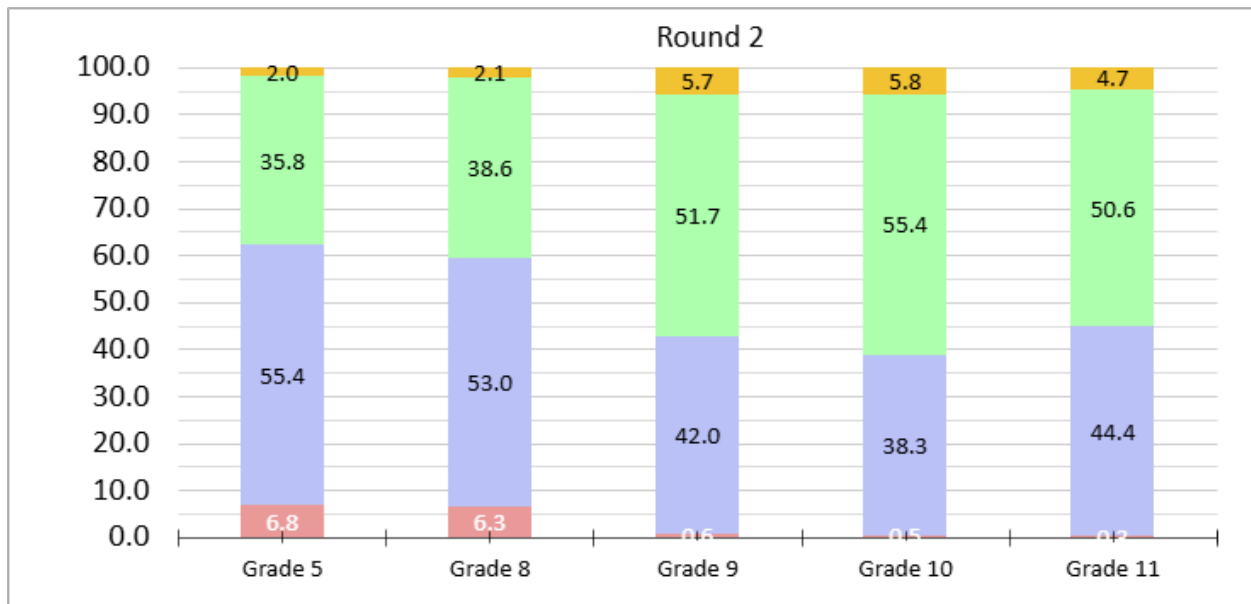


Figure 27. Science Impact Data: Round 2

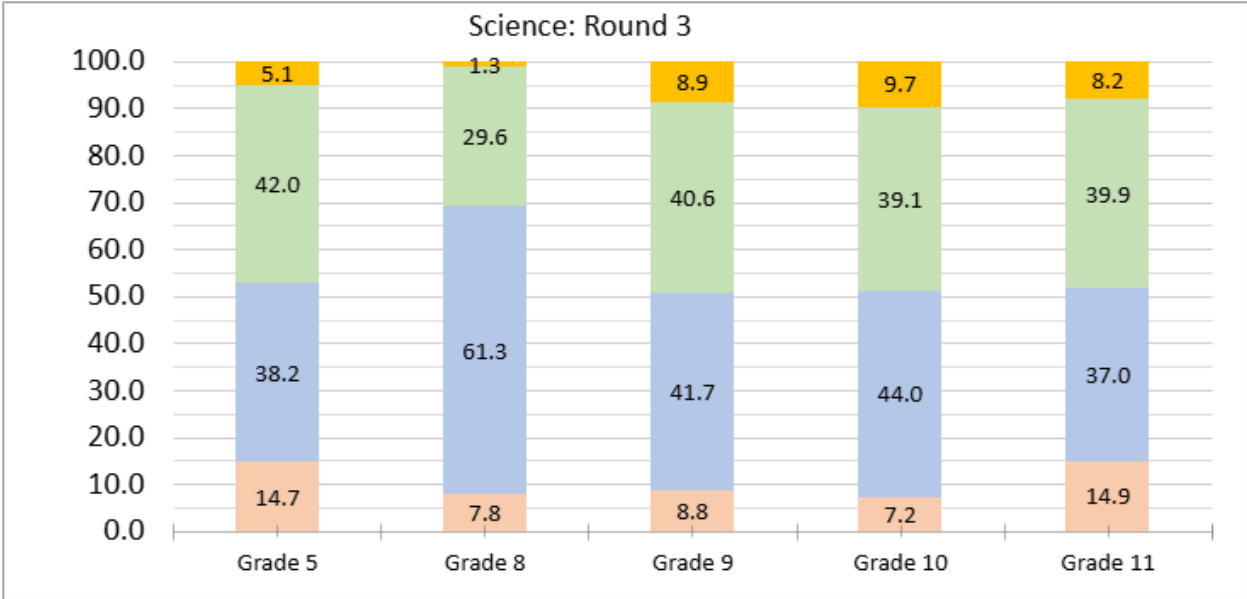


Figure 28. Science Impact Data Round 3

ACT Results for High School Grades

Rounds 1 and 2 used the Modified Angoff Yes/No approach, and Rounds 3 and 4 used the empirical approach. Table 42 to Table 45 contain the cut score statistics from each round on the 1–36 score scale. The Round 1 cut scores for Proficient were all higher than the ACT Benchmarks, with the exception of Writing, which does not have a Benchmark. The cut scores for Above Proficient were also exceptionally high, with the exception of the cut score for Writing, and the cut scores for Science and English were set at the maximum score of 36. The cut scores for Approaching Proficient were near chance for Math (13) and English (12). Panelists were informed that a score of 12 can be considered a “chance” score, or what a student might earn if they answered all items randomly on the assessment.

Table 42. Round 1 Cut Score Statistics

	Level	Panelists	Min	Q1	Median	Mean	Q3	Max
Math	Above	14	30	31	32	33	35	36
	Proficient	14	17	21	24	24	26	30
	Approaching	14	6	13	13	14	16	20
Science	Above	10	27	32	36	33	36	36
	Proficient	10	19	21	25	25	28	36
	Approaching	10	8	10	14	14	17	24
English	Above	9	32	36	36	35	36	36
	Proficient	9	15	19	23	24	27	36
	Approaching	9	4	5	12	13	16	26
Reading	Above	11	26	32	35	33	36	36
	Proficient	11	18	22	24	24	26	32
	Approaching	11	7	12	14	14	16	18
Writing	Above	10	23	24	25	26	26	31
	Proficient	10	18	20	20	20	22	23
	Approaching	10	12	14	15	15	17	18

For Round 2, panelists generally adjusted their cuts downward, and in some cases overcorrected, particularly for the Approaching Proficient cut, for which Science, English, and Reading had median cut scores below chance (7, 9, and 9, respectively). The Round 2 cut scores for Writing were largely the same as Round 1, with the exception of the Approaching Proficient cut score dropping by one point on the 1–36 scale.

Table 43. Round 2 Cut Score Statistics

	Level	Panelists	Min	Q1	Median	Mean	Q3	Max
Math	Above	14	28	29	30	30	31	32
	Proficient	14	16	17	19	19	20	25
	Approaching	14	9	11	12	12	13	17
Science	Above	10	28	30	31	30	31	32
	Proficient	10	17	17	18	19	20	23
	Approaching	10	3	3	7	6	8	11
English	Above	9	21	25	29	29	32	36
	Proficient	9	11	14	15	19	21	35
	Approaching	9	2	5	9	9	11	23
Reading	Above	11	11	30	32	30	34	36
	Proficient	11	6	13	19	17	22	28
	Approaching	11	1	6	9	8	10	14
Writing	Above	10	23	23	25	25	25	27
	Proficient	10	17	18	20	20	20	23
	Approaching	10	9	12	14	14	16	18

Round 3 was the first round of cut scores after transitioning to the empirical approach, and the English, Reading, and Writing panels combined into a single ELA group. The Round 3 cut scores in for Proficient in Math and Science were the same as Round 2, and the average of the English, Reading, and Writing Proficient cut scores from Round 2 was an 18, which is same as the Round 3 ELA Proficient cut score. The Above Proficient cut scores dropped across subject areas, and the Approaching Proficient cut scores increased across subject areas (the ELA cut scores based on the Round 2 cuts in English, Reading, and Writing were 29 for Above Proficient and 11 for Approaching Proficient).

Table 44. Round 3 Cut Score Statistics

	Level	Panelists	Min	Q1	Median	Mean	Q3	Max
Math	Above	14	21	25	28	27	29	29
	Proficient	14	16	18	19	18	19	19
	Approaching	14	14	14	15	15	15	16
Science	Above	11	25	26	27	27	28	28
	Proficient	11	18	18	18	19	18	23
	Approaching	11	12	14	15	14	15	15
ELA	Above	14	22	23	27	26	28	31
	Proficient	14	17	18	18	19	21	23
	Approaching	14	13	13	14	14	15	16

The Round 4 cut scores were largely the same as the Round 3 cut scores, with minor exceptions. The Proficient cut score for Science increased by one point, the Approaching

Proficient cut score for Science decreased by one point, and the Approaching Proficient cut score for ELA decreased by one point. Consensus was not reached for any of the cut scores, but the interquartile range tended to within one score point and the full range tended to be within 2–3 points with the exception of the Above Proficient cut score for ELA, which had a range of 5 points.

Table 45. Round 4 Cut Score Statistics

	Level	Panelists	Min	Q1	Median	Mean	Q3	Max
Math	Above	14	25	27	28	27	28	28
	Proficient	14	18	18	19	19	19	19
	Approaching	14	14	15	15	15	15	16
Science	Above	11	25	27	27	27	27	28
	Proficient	11	18	19	19	19	19	20
	Approaching	11	14	14	14	14	15	15
ELA	Above	14	24	26	27	26	27	29
	Proficient	14	18	18	18	18	18	19
	Approaching	14	13	13	13	13	13	14

Figure 29 to Figure 31 contain the Round 4 dot plots for each subject. These plots show the spread of the final cut scores from each individual panelist.

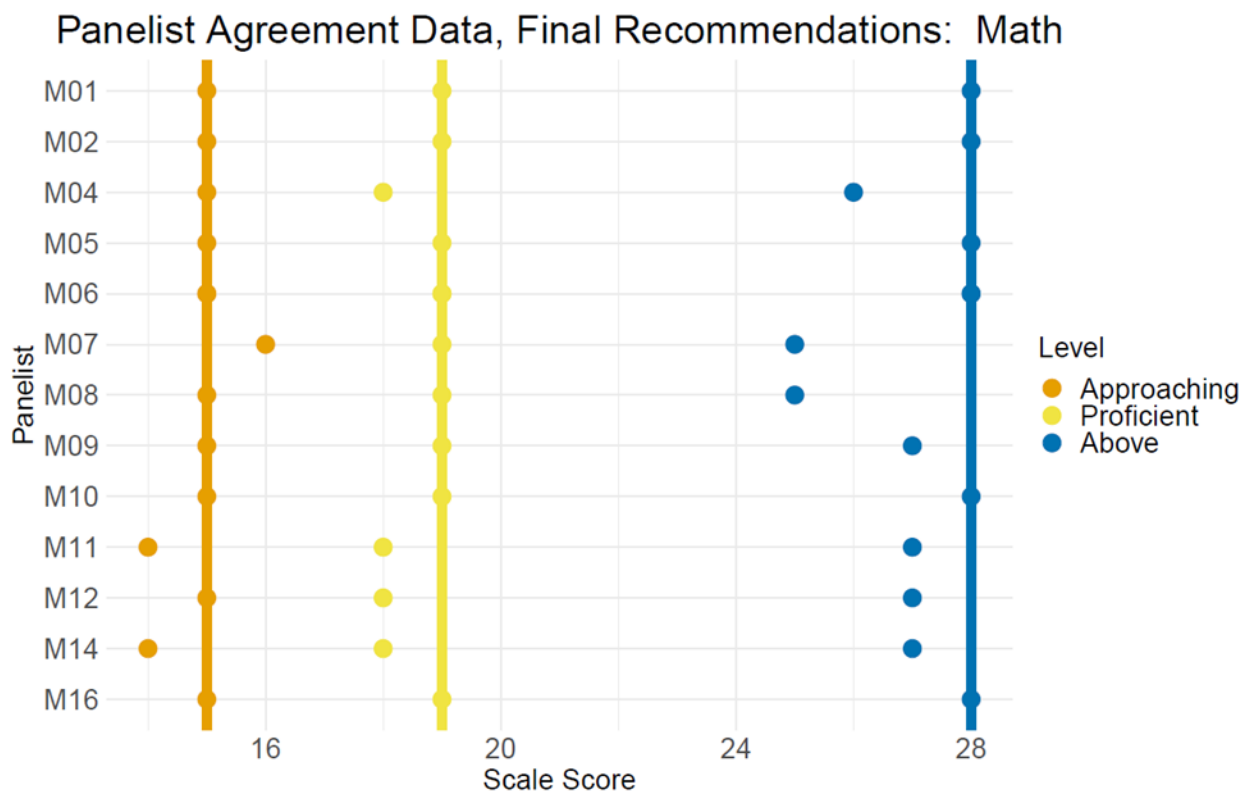


Figure 29. Panelist Agreement: Math

Panelist Agreement Data, Final Recommendations: Science

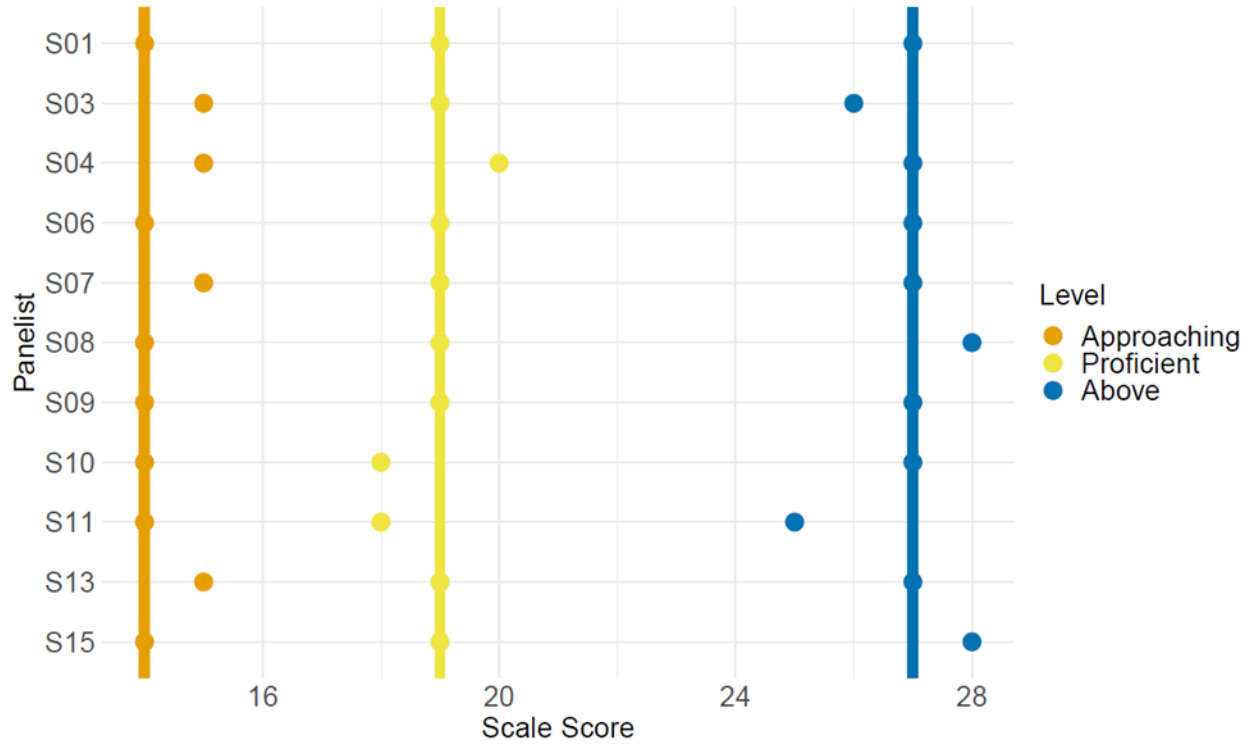


Figure 30. Panelist Agreement: Science

Panelist Agreement Data, Final Recommendations: ELA

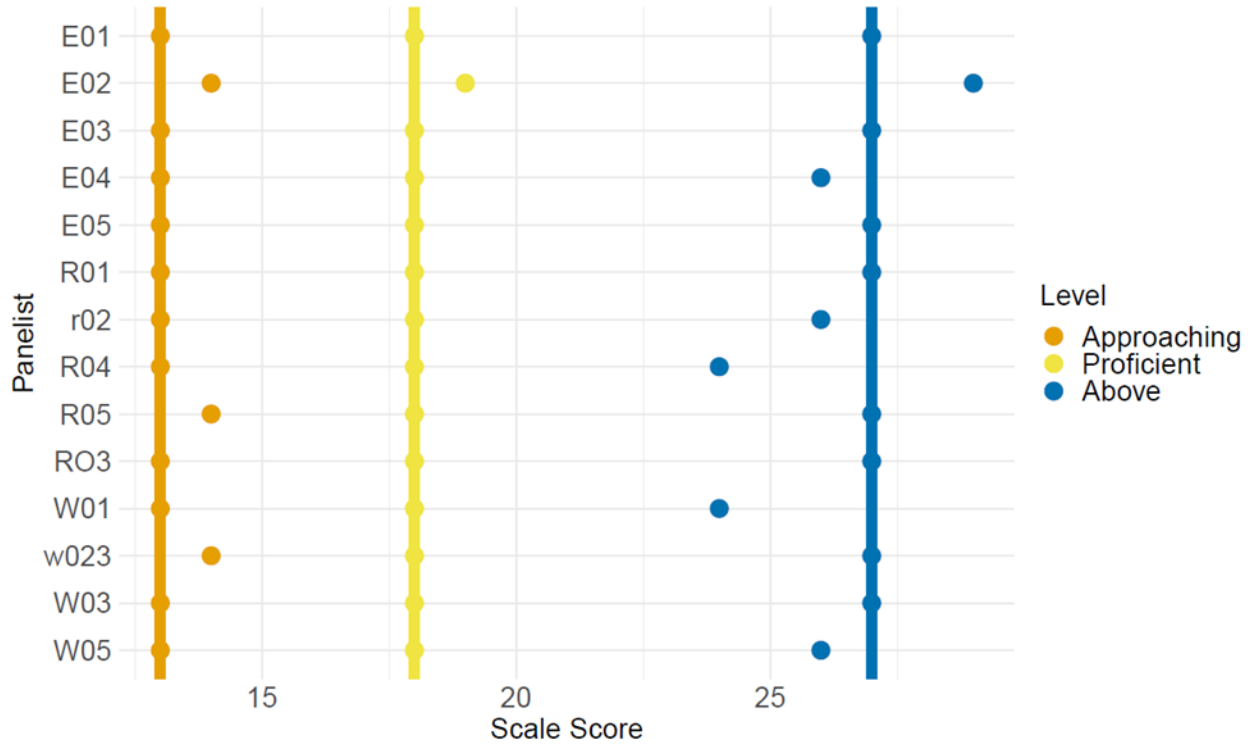


Figure 31. Panelist Agreement: ELA

Figure 32 contains the Round 4 impact data. Across subject areas, approximately 10% of students scored in the Above Proficient performance level, with 10% in Math, 8% in Science, and 11% in ELA. For Science and ELA, approximately 40% of students scored in the Proficient performance level (40% and 41%, respectively), while 29% scored at proficient in math. Similar percentages of students scored at Approaching Proficient in Math (36%) and Science (37%), while a smaller percentage of students scored at Approaching Proficient in ELA (28%), and the percentages of students scoring Below Proficient ranged from 25% in Math to 15% in Science, with 20% scoring Below Proficient in ELA.

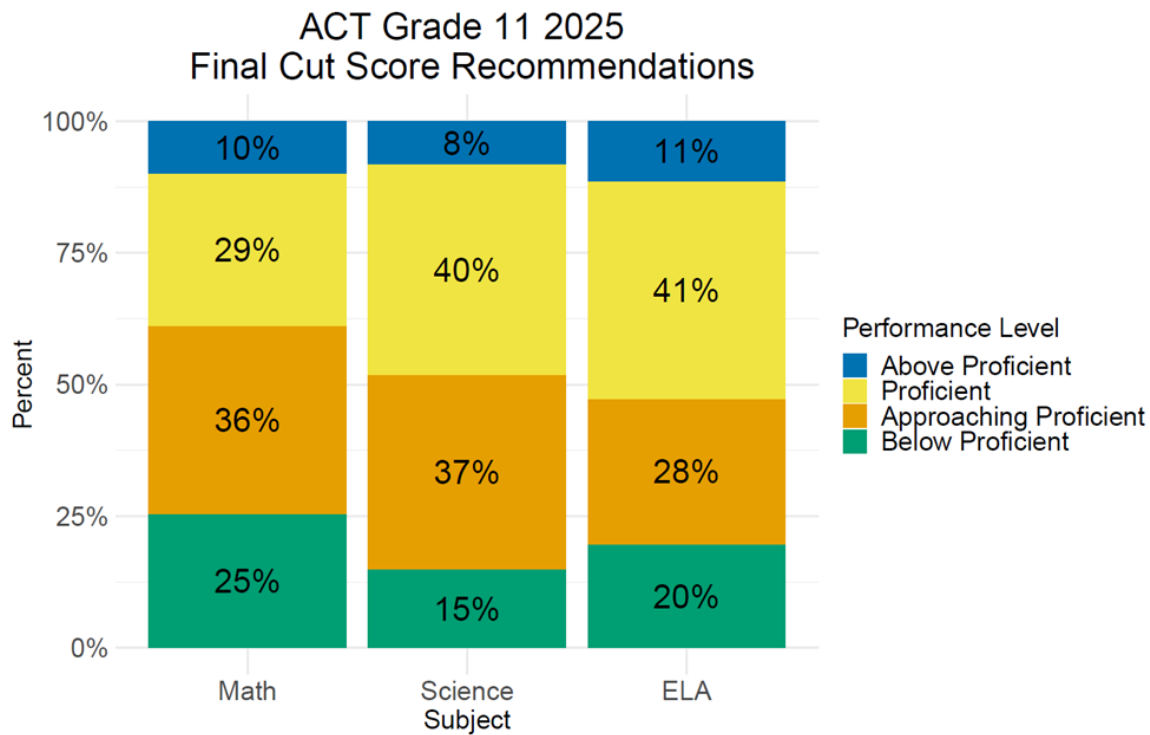


Figure 32. Round 4 Impact Data: Percentage of Illinois Students Scoring Within Each Performance Level

Table 46 contains the PreACT Secure and PreACT 9 Secure scores linked to the Round 4 ACT cut scores using an equipercentile linking. The percent of students who scored at or above each cut score are presented; the percentages of students scoring at or above each PreACT Secure cut score are slightly different due to the score scale and different percentages of students scoring at or above a specific score point. For example, the Approaching cut score for ACT math was 15, and 75% of students scored at or above 15. The 75th percentile for PreACT Secure math in grade 10 was a score of 15, but due to the score scale of 1–35 for PreACT Secure math, several percentiles corresponded to a score of 15, and in total, 79% of students scored at or above 15 on PreACT Secure math in grade 10. Similarly, the 75th percentile for PreACT 9 Secure math was a score of 14, but the total percentage of students scoring at or above 14 on PreACT 9 Secure math was 93% because a score of 14 (on a 1–32 scale) corresponded to multiple percentiles.

Table 46. PreACT Secure Scores Linked to the Round 4 ACT Cut Scores

Subject	Level	ACT Score	PreACT Secure Score	PreACT 9 Secure Score	Percent At/Above ACT	Percent At/Above PreACT Secure	Percent At/Above PreACT 9 Secure
Math	Above	28	25	24	10	11	13
Math	Proficient	19	17	16	39	42	47
Math	Approaching	15	15	14	75	79	93
Science	Above	27	25	25	8	10	9
Science	Proficient	19	16	14	48	49	49
Science	Approaching	14	11	11	85	93	91
ELA	Above	27	26	25	11	13	11
ELA	Proficient	18	15	14	53	54	61
ELA	Approaching	13	11	11	80	91	89

Vertical Articulation Results

The vertical articulation process was a critical final step in the standard setting, intended to ensure that performance standards carry coherent meaning as students progress through grades. By design, performance level definitions (PLDs) were intended to scale up in rigor across grades, and the cut scores should reflect those increasing expectations. Here we reflect on how the vertical articulation achieved coherence both within content areas and, to the extent applicable, across content areas.

Within-Content Vertical Articulation

The vertical articulation process played a pivotal role in the standard setting meeting, where panelists aimed to ensure that proficiency expectations were consistently aligned across grades and subjects, specifically in English/Language Arts (ELA), Mathematics, and Science for students in Grades 3 through 11. The primary purpose of this process was to establish fair, reliable, and coherent academic benchmarks that accurately reflected student readiness for college and career. This initiative addressed longstanding issues with misaligned proficiency cut scores by evaluating patterns of results across grades and making necessary adjustments to cut scores for performance levels.

During the vertical articulation meeting, panelists engaged in several key activities to achieve these objectives. Panelists began with introductions and a discussion on the purpose of vertical articulation, setting the stage for a collaborative effort. Panelists then reviewed cross-grade performance level descriptors (PLDs), which outlined the specific expectations of student performance at each level. This review allowed them to understand the progression of knowledge and skills across performance levels within each subject.

Next, panelists examined cross-grade impact data, which provided insights into the percentage of students classified into each performance level based on the recommended cut scores. This data served as a reality check, helping us assess whether the current cut score recommendations aligned with expected student performance patterns across grades.

Following the data review, panelists participated in breakout room discussions, where panelists evaluated the reasonableness of the recommended cut scores and considered potential changes. These discussions were guided by questions about the differences in expectations across grades, the factors supporting these expectations, and whether the proposed changes were supported by the PLDs.

Vertical Articulation Adjustments

After reviewing the cuts across grades, the Vertical Articulation committees made minor adjustments to ensure a coherent progression. The tables below highlight any adjustments made during articulation in bold in the Articulation Adjustment column.

Table 47. Vertical Articulation Adjustments to Cut Scores: ELA

Subject	Grade	Performance Level	Final Round SS Cuts	Articulation Adjustments
ELA	3	Above Proficient	788	780
		Proficient	749	735
		Approaching Proficient	685	685
	4	Above Proficient	797	780
		Proficient	752	737
		Approaching Proficient	675	695
	5	Above Proficient	795	780
		Proficient	744	739
		Approaching Proficient	700	700
	6	Above Proficient	773	780
		Proficient	739	741
		Approaching Proficient	702	705
	7	Above Proficient	783	785
		Proficient	746	743
		Approaching Proficient	716	710
	8	Above Proficient	802	795
		Proficient	758	745
		Approaching Proficient	704	710
	9	Above Proficient	25	25
		Proficient	14	14
		Approaching Proficient	11	11
	10	Above Proficient	26	26
		Proficient	15	15
		Approaching Proficient	11	12
	11	Above Proficient	27	27
		Proficient	18	18
		Approaching Proficient	13	13

ELA Impact Data

Figure 33 and Figure 34 display the impact of the cut scores on student performance in ELA following vertical articulation. Figure 34 summarizes these performance levels into two main categories: Proficient and Above, and Non-Proficient. The percentage of students at or above Proficient ranged from 46.1% in Grade 3 to 60.8% in Grade 9, with most grades

showing more than half of students reaching proficiency. The panelists explained that they felt that the consistency across grades supported the reasonableness of the ELA cuts.

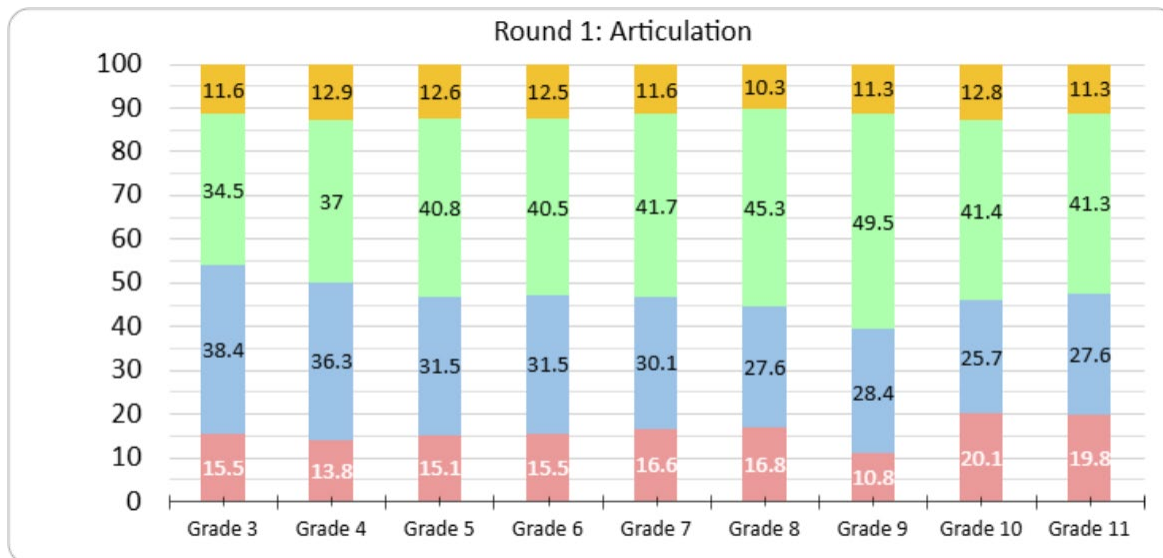


Figure 33. ELA Impact Data by Performance Level After Vertical Articulation

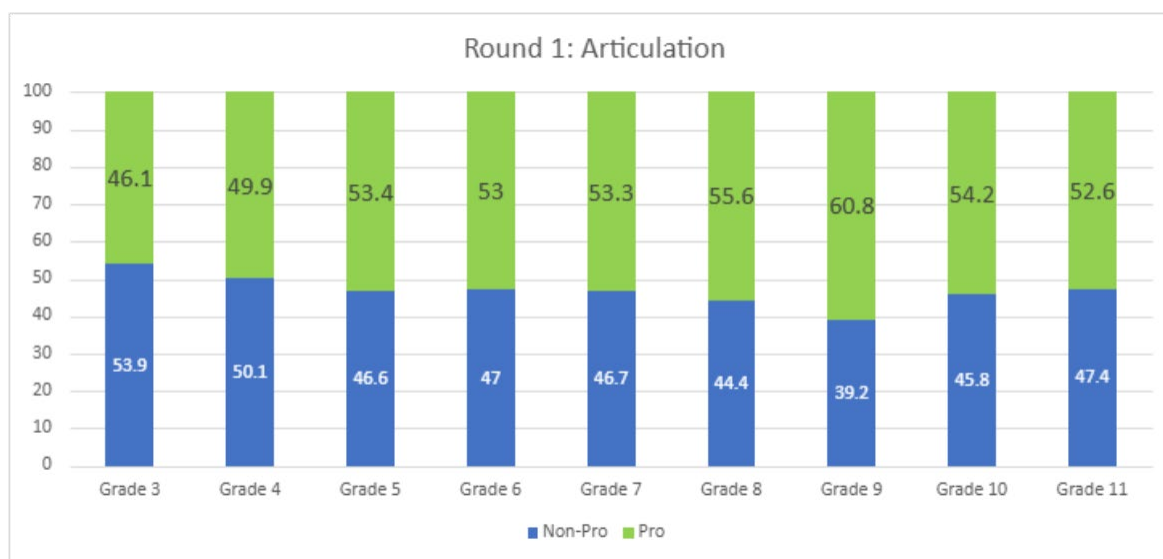


Figure 34. ELA Impact Data by Proficient vs. Non-Proficient After Vertical Articulation

After reviewing the cuts across grades, the Math Vertical Articulation committees made minor adjustments to ensure a coherent progression. The tables below highlight any adjustments made during articulation in bold in the Articulation Adjustment column.

Table 48. Vertical Articulation Adjustments to Cut Scores: Mathematics

Subject	Grade	Performance Level	Final SS	Articulation Adjustments
Math	3	Above Proficient	771	781
		Proficient	732	732
		Approaching Proficient	700	705
	4	Above Proficient	784	784
		Proficient	740	740
		Approaching Proficient	708	708
	5	Above Proficient	782	782
		Proficient	748	740
		Approaching Proficient	709	709
	6	Above Proficient	773	773
		Proficient	742	742
		Approaching Proficient	702	705
	7	Above Proficient	780	781
		Proficient	745	745
		Approaching Proficient	712	712
	8	Above Proficient	791	791
		Proficient	746	745
		Approaching Proficient	684	705
	9	Above Proficient	24	24
		Proficient	16	17
		Approaching Proficient	14	15
	10	Above Proficient	25	25
		Proficient	17	18
		Approaching Proficient	15	15
	11	Above Proficient	28	28
		Proficient	19	19
		Approaching Proficient	15	15

Mathematics Impact Data

Figure 35 and Figure 36 show the distribution of student performance in Mathematics. Figure 36 presents the percentage of students who were Proficient or Above versus those who were not. The percentage of students meeting or exceeding proficiency ranged from 32% in Grade 6 to 48.8% in Grade 3, with most grades in the 30s or low 40s. The panelists discussed that they felt that this impact distribution confirmed that the math cuts created meaningful distinctions in student performance while ensuring grade-to-grade coherence.

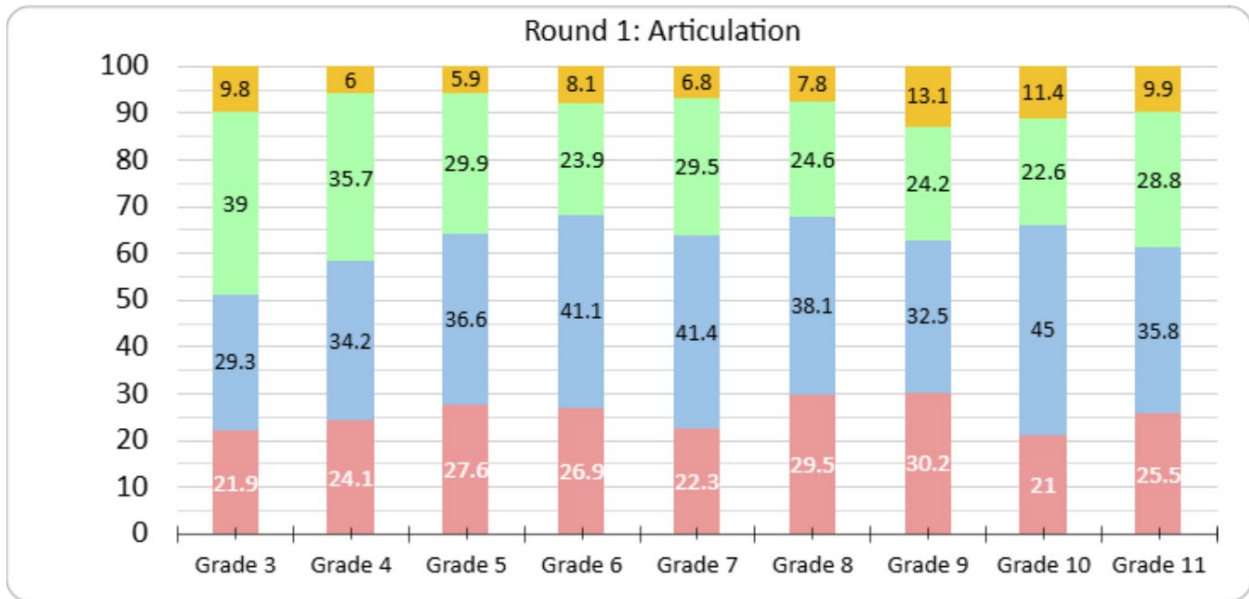


Figure 35. Mathematics Impact Data by Performance Level After Vertical Articulation

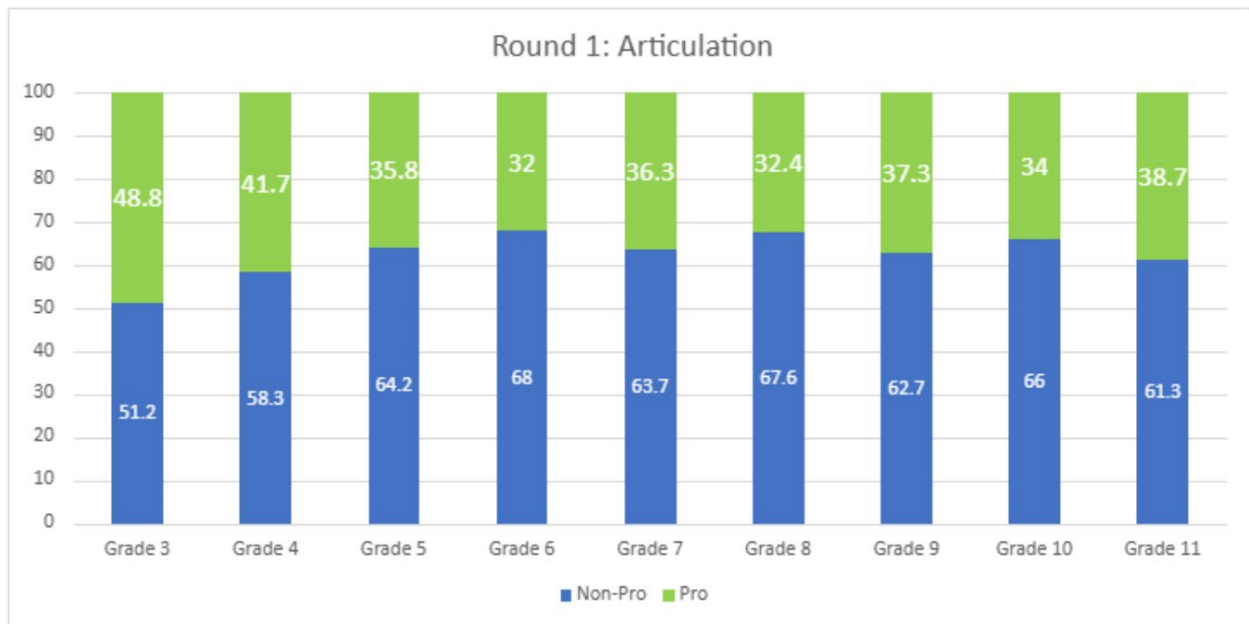


Figure 36. Mathematics Impact Data by Proficient vs. Non-Proficient After Vertical Articulation

After reviewing the cuts across grades, the Science Vertical Articulation committees made minor adjustments to ensure a coherent progression. The table below highlights any adjustments made during articulation in bold in the Articulation Adjustment column.

Table 49. Vertical Articulation Adjustments to Cut Scores: Science

Subject	Grade	Performance Level	Final SS	Articulation Adjustments
Science	5	Above Proficient	856	856
		Proficient	804	812
		Approaching Proficient	766	770
	8	Above Proficient	876	856
		Proficient	822	812
		Approaching Proficient	760	770
	9	Above Proficient	25	25
		Proficient	14	14
		Approaching Proficient	11	11
10	Above Proficient	25	25	
	Proficient	16	16	
	Approaching Proficient	11	11	
11	Above Proficient	27	27	
	Proficient	19	19	
	Approaching Proficient	14	14	

Science Impact Data

Figure 37 and Figure 38 depict the Science impact data after vertical articulation. Figure 38 illustrates the coherence of proficiency rates across grades. The percentage of students scoring Proficient or Above ranged from 37.7% in Grade 5 to 49.5% in Grade 9, with most grades clustering in the 40s. The relatively even performance distribution provided assurance to the panelists that the Science cut scores produced consistent results across the grade bands.

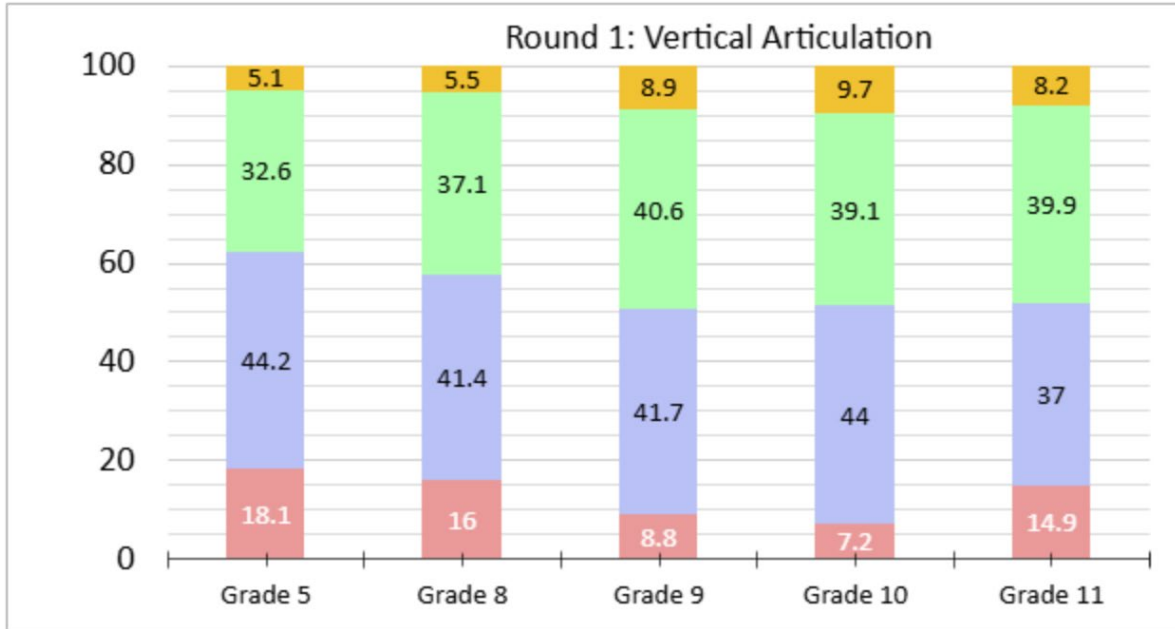


Figure 37. Science Impact Data by Performance Level After Vertical Articulation

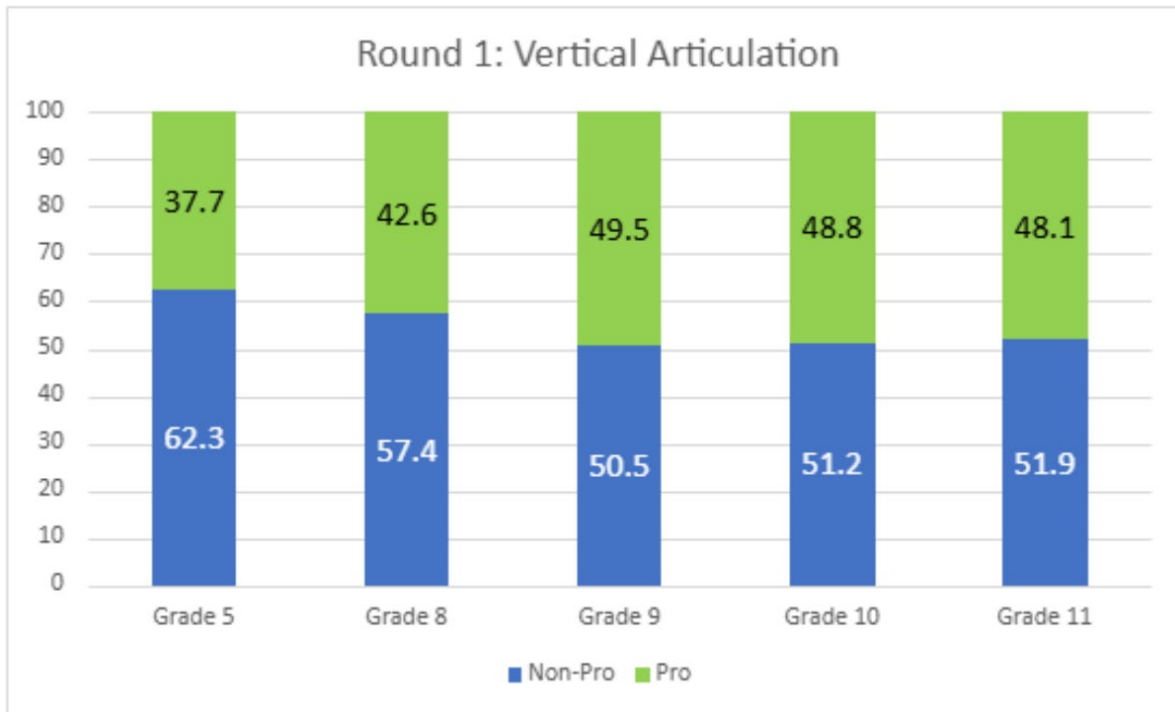


Figure 38. Science Impact Data by Proficient vs. Non-Proficient After Vertical Articulation

Cross-Content Coherence

After the Vertical Articulation processes concluded, all committees came back together in a general session. In the general session, panelists engaged in a cross-content area impact data reasonableness discussion, where they collectively assessed the consistency of expectations and results across different subjects. Finally, the Content Coherence facilitator wrapped up the meeting by evaluating the outcomes and requesting any final adjustments to the cut scores, ensuring that they fell within the recommended ranges and were supported by the content basis.

A joint review examined whether “Proficient” signifies comparable rigor across ELA, Mathematics, and Science. Panels confirmed that the resulting cuts and PLDs tell a coherent cross-content story; no further changes were required following the content-coherence discussion. They did vote on all the cutscores using an online tool that captured their ratings for the following statements:

1. I support the cutscores.
2. I can live with the cutscores.
3. I reject the cutscores.

Panelists had to enter rating for each set of grade/subject/cutscore combination. All ratings either supported the cutscores or could live with the cutscores.

Overall, the goal of the vertical articulation and content coherence review was to establish unified academic standards that were reasonable and equitable for students throughout Illinois, addressing previous issues with misaligned proficiency cut scores and setting a solid foundation for future assessments. These recommended cut scores were approved by the Illinois State Board of Education (ISBE) with no changes.

Figure 39 compares the percentage of students at the Proficient level by subject and grade. ELA is in yellow, Math is in blue, and Science is in green. While there is some natural variation across content areas, ELA generally had higher proficiency rates in the early grades, with Math showing the lowest rates in upper middle grades. Science proficiency rates were in the middle range and increased in the ACT grades. Notably, Grade 9 shows the greatest divergence, with ELA proficiency at 60.8%, Math at 37.3%, and Science at 49.5%. Still, across all subjects, the articulation process led to performance distributions that panelists considered reasonable, fair, and aligned to expectations for student readiness.

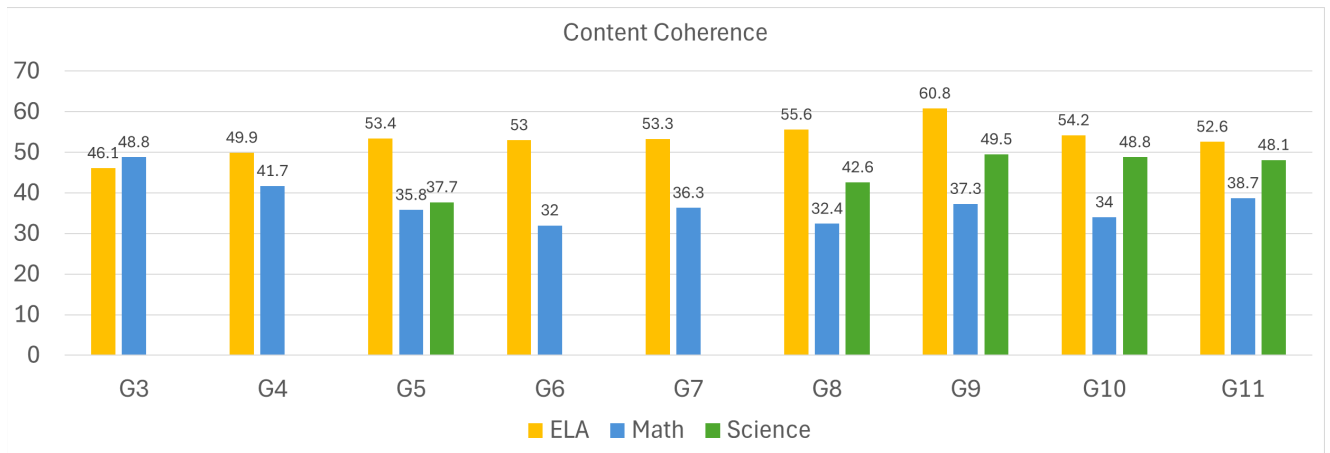


Figure 39. Percentage of Proficient Students in each Grade by Subject

In conclusion, the impact review supported the soundness of the recommended cut scores. Panelists across ELA, Math, and Science indicated that the final performance distributions demonstrated coherence across grades and content areas. After thorough consideration, the panels unanimously agreed that the cuts were reasonable and chose to retain the established cut scores without further modification.

Evaluation Survey and Panelist Feedback

The Vertical Articulation Evaluation for the Illinois Assessment of Readiness (IAR), Illinois Science Assessment (ISA), and ACT Assessments was conducted on July 18, 2025, to gather feedback from participants in the vertical articulation meeting, including the final content reasonableness review across content areas. The evaluation aimed to assess the effectiveness of the vertical articulation process, which focused on ensuring cut score coherence across grade levels for ELA (Grades 3-11), Mathematics (Grades 3-11), and Science (Grades 5, 8, and 9-11). The evaluation collected participant feedback to refine the process, enhance alignment with student success, and provide validity evidence for the standard setting process.

The evaluation consisted of two main sections: a survey assessing the success of meeting components and the adequacy of training elements, and questions gauging participants' understanding and confidence in the recommended cut scores. The survey included multiple-choice questions with options ranging from "Not Successful" to "Very Successful" for meeting components, and from "Not Adequate" to "More than Adequate" for training elements. Participants were also asked to indicate their confidence in the final cut score recommendations for each subject area (ELA, Mathematics, and Science) across performance levels (Approaching Proficient, Proficient, and Above Proficient). Open-

ended questions allowed participants to elaborate on concerns about cut score recommendations and provide suggestions for improving the vertical articulation process.

The evaluation results were intended to help ISBE and Pearson improve future vertical articulation processes by identifying strengths and areas for improvement in training, collaboration, and the overall standard setting process. The feedback also served as a critical component for ensuring the coherence and validity of the cut score recommendations across grade levels and content areas. See Appendix G for a copy of the survey questions.

The Vertical Articulation Evaluation surveys for the Illinois Assessment of Readiness (IAR), Illinois Science Assessment (ISA), and ACT assessments focused on assessing the effectiveness of the vertical articulation process, evaluating the success of meeting components, gauging confidence in final cut score recommendations, and identifying strengths and areas for improvement. Here's a summarized overview:

Success of Meeting Components:

- **Introduction to Vertical Articulation Process:** Generally rated as "Very Successful," with average scores around 3.6 to 3.8. Most participants found the introduction effective in setting the stage for the process.
- **Review of the PLDs: Rated as "Very Successful,"** with scores from 3.7 to 3.8. Participants appreciated the thorough review of Performance Level Descriptors.
- **Review of Cross-Grade Impact Data:** Seen as "Successful" to "Very Successful," with scores from 3.5 to 3.9. The data review was considered crucial for understanding student performance across grades.
- **Use of Interactive Vertical Articulation Spreadsheet:** Rated as "Successful" to "Very Successful," with scores from 3.7 to 3.9. The tool was valued for facilitating data analysis and discussions.
- **Discussion of Recommended Changes Across Grades:** Generally "Very Successful," with scores from 3.7 to 3.8. Participants felt the discussions were productive and led to meaningful recommendations.
- **General Session Review and Discussion:** Rated as "Successful" to "Very Successful," with scores from 3.5 to 3.7. The session was effective in reviewing content reasonableness across areas.

Understanding and Support for Cut Scores:

- **Understanding of Process:** Nearly all participants (100% in some groups) understood the process for recommending changes to scaled score cuts.
- **Support for Cut Scores:** Participants unanimously supported the cut scores recommended after the vertical articulation process and the final cross-content area reasonableness review.

Adequacy of Vertical Articulation Training:

- **Amount of Time Spent Reviewing PLDs:** Rated as "Adequate" to "More Than Adequate," with scores from 3.3 to 3.6.
- **Amount of Time Discussing Impact Data Across Grades:** Generally "Adequate" to "More Than Adequate," with scores from 3.5 to 3.7.
- **Amount of Time Working with Interactive Spreadsheet:** Rated as "Adequate" to "More Than Adequate," with scores from 3.4 to 3.7.
- **Amount of Time for Content Reasonableness Review:** Seen as "Adequate" to "More Than Adequate," with scores from 3.5 to 3.6.

Confidence in Final Cut Score Recommendations:

- ELA, Math, and Science: Participants expressed confidence in the final cut score recommendations, with most feeling "Confident" or "Approaching Proficient." Scores for confidence ranged from 2.6 to 3.0 across different proficiency levels.

Additional Feedback:

- Participants appreciated the collaborative nature of the process and the opportunity to contribute to important decisions affecting student assessments.
- Some suggestions included involving more teachers for broader representation and providing more time for discussions within grade bands.
- Concerns were raised about the potential disconnect between proficiency levels in Math and ELA, with some participants worried about the implications for district focus and support.
- Overall, the Vertical Articulation Evaluation surveys indicate a successful process with strong support for the cut scores and confidence in the recommendations. Participants valued the opportunity to engage in meaningful discussions and contribute to the alignment of assessments across grade levels.

Approval of Standard Setting Results

On August 13, 2025, the Illinois State Board of Education (ISBE) approved new, research-informed proficiency benchmarks for state assessments in English Language Arts, mathematics, and science. This decision followed an extensive 18-month statewide engagement and design process that aimed to align performance levels with actual college and career readiness expectations. The initiative engaged over 100 educators and specialists who developed performance level descriptors, and 147 teachers who evaluated item difficulty to recommend cut scores. The process also included analysis of college placement data and student success probabilities, with oversight from the Technical Advisory Committee and validation by the Center for Assessment. The newly adopted benchmarks unified performance levels across all assessments and recalibrated proficiency rates to establish a new baseline for longitudinal tracking.

During the Board meeting, State Superintendent Dr. Tony Sanders emphasized that the new benchmarks corrected years of mislabeling students and better reflected their readiness for college and career pathways. Board members expressed strong support for the changes, highlighting the importance of accurate data to ensure equitable access to advanced coursework and postsecondary opportunities. The Board also acknowledged the collaborative nature of the process, which included input from students, educators, postsecondary institutions, and community members. The approval marked a significant milestone in Illinois' commitment to transparency, equity, and continuous improvement in educational assessment.

The new cut score and cut score ranges for the IAR and ISA are presented in Table 50 and the new cuts and ranges for the ACT are presented in Table 51. Since IAR and ISA are calibrated using Item Response Theory (IRT), contains that cut scores on the theta metric, along with the slope and intercept used to calculate the scaled scores on the theta to scale score transformation.

Table 50. Adopted Cutscores and Cutscore Ranges for IAR and ISA

Subject	GR	Below Proficient Cut	Below Proficient Range	Approaching Proficient Cut	Approaching Proficient Range	Proficient Cut	Proficient Range	Above Proficient Cut	Above Proficient Range
ELA	3	650	650-684	685	685-734	735	735-779	780	780-850
ELA	4	650	650-694	695	695-736	737	737-779	780	780-850
ELA	5	650	650-699	700	700-738	739	739-779	780	780-850
ELA	6	650	650-704	705	705-740	741	741-779	780	780-850
ELA	7	650	650-709	710	710-742	743	743-784	785	785-850
ELA	8	650	650-709	710	710-744	745	745-794	795	795-850
Math	3	650	650-704	705	705-731	732	732-780	781	781-850
Math	4	650	650-707	708	708-739	740	740-783	784	784-850
Math	5	650	650-708	709	709-739	740	740-781	782	782-850
Math	6	650	650-704	705	705-741	742	742-772	773	773-850
Math	7	650	650-711	712	712-744	745	745-780	781	781-850
Math	8	650	650-704	705	705-744	745	745-790	791	791-850
SCI	5	700	700-769	770	770-811	812	812-855	856	856-900
SCI	8	700	700-769	770	770-811	812	812-855	856	856-900

Table 51. Adopted Cutscores and Cutscore Ranges for ACT

Subject	GR	Below Proficient Cut	Below Proficient Range	Approaching Proficient Cut	Approaching Proficient Range	Proficient Cut	Proficient Range	Above Proficient Cut	Above Proficient Range
ELA	9	1	1-10	11	11-13	14	14-24	25	25-32
ELA	10	1	1-11	12	12-14	15	15-25	26	26-35
ELA	11	1	1-12	13	13-17	18	18-26	27	27-36
Math	9	1	1-14	15	15-16	17	14-23	24	24-32
Math	10	1	1-14	15	15-17	18	18-24	25	25-35
Math	11	1	1-14	15	15-18	19	19-27	28	28-36
SCI	9	1	1-10	11	11-13	14	14-24	25	25-32
SCI	10	1	1-10	11	11-15	16	16-24	25	25-35
SCI	11	1	1-13	14	14-18	19	19-26	27	27-36

Table 52. Scale Transformation Constants For Theta to Scaled Score Transformations

Subject	Grade	SS Cut1	SS Cut2	SS Cut3	LOSS	HOSS	Scaling Constant: Slope	Scaling Constant: Intercept	Theta Cut1	Theta Cut2	Theta Cut3
ELA	3	685	735	780	650	850	36.7227	735.4297	-1.3733	-0.0117	1.2137
ELA	4	695	737	780	650	850	31.5462	741.0214	-1.4589	-0.1275	1.2356
ELA	5	700	739	780	650	850	29.458	739.505	-1.3411	-0.0171	1.3747
ELA	6	705	741	780	650	850	28.316	738.6673	-1.1890	0.0824	1.4597
ELA	7	710	743	785	650	850	33.9161	742.3542	-0.9539	0.0190	1.2574
ELA	8	710	745	795	650	850	34.1183	743.433	-0.9799	0.0459	1.5114
Math	3	705	732	781	650	850	32.1135	745.4119	-1.2584	-0.4176	1.1082
Math	4	708	740	784	650	850	29.9167	741.4049	-1.1166	-0.0470	1.4238
Math	5	709	740	782	650	850	29.0301	742.2997	-1.1471	-0.0792	1.3676
Math	6	705	742	773	650	850	28.1465	738.9252	-1.2053	0.1092	1.2106
Math	7	712	745	781	650	850	25.1033	736.3102	-0.9684	0.3462	1.7802
Math	8	705	745	791	650	850	32.9505	729.164	-0.7333	0.4806	1.8766
SCI	5	770	812	856	700	900	27.7915	783.7664	-0.4953	1.0159	2.5991
SCI	8	770	812	856	700	900	29.2258	792.9673	-0.7859	0.6512	2.1567

References

- Allen, J. (2013). *Updating the ACT College Readiness Benchmarks*. (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc. Retrieved from <http://www.act.org/content/act/en/research/pdfs/updating-the-actcollegereadinessbenchmarks.html>
- Allen, J., & Sconing, J. (2005). *Using ACT Assessment scores to set benchmarks for college readiness*. (ACT Research Report No. 2005-3). Iowa City, IA: ACT, Inc. Retrieved from <http://www.act.org/content/act/en/research/pdfs/using-act-assessmentstosetbenchmarksforcollegereadiness.html>
- Camara, W. J., Allen, J. M., & Moore, J. L. (2017). Empirically-based college and career readiness cut scores and performance standards. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice*. New York, NY: Routledge.
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2022). *shiny: Web Application Framework for R*. R package version 1.7.4, <https://github.com/rstudio/shiny>, <https://shiny.posit.co/>
- Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23(4), 31-50.
- Davis, L. L., & Moyer, E. E. (2015). Yes/No Angoff variants in practice: Design choices and consequences. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed., pp. 61-78). New York, NY: Routledge.
- Davis, L. L., & Moyer, E. L. (2015). *PARCC performance level setting technical report*. Partnership for Assessment of Readiness for College and Careers (PARCC).
- Haertel, E. H., Beimers, J. N., & Miles, J. A. (2012). The Briefing Book method. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd ed., pp. 283-299). New York, NY: Routledge.
- Haertel, E., Buckendahl, C., & Schneider, M. (2012). Briefing books and validity evidence in standard setting. In G. J. Cizek (Ed.), *Setting performance standards* (2nd ed., pp. 307-332). New York, NY: Routledge.
- Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 41-55.

- Illinois Standard Setting Plan. (Version 1.4). *Standard setting plan for Illinois assessments* (Drafted by Pearson & ACT, April 15, 2025).
- Illinois State Board of Education. (2025). *Illinois IAR, ISA, and ACT 2025 standard setting executive summary*. (Prepared by Pearson and ACT, July 2025.)
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed). New York, NY: Springer.
- Noble, J., & Radunzel, J. (2012). *Predicting long-term college success through degree completion using ACT® Composite score, ACT Benchmarks, and high school grade point average*. ACT Research Report 2012-5.
- Plake, B. S., & Cizek, G. J. (2012). Variations on a theme: The Modified Angoff, Extended Angoff, and Yes/No standard setting methods. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd ed., pp. 79–106). New York, NY: Routledge.
- Plake, B. S., Ferdous, A. A., Impara, J. C., & Buckendahl, C. W. (2005). *Setting multiple performance standards using the Yes/No method: An alternative item mapping method*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
-

Appendices

Appendix A: Agendas

1. Panelist Agenda for Grades 4, 6, 8 – ELA, Math, and Science
2. Panelist Agenda for Grades 3, 5, 7 – ELA & Math
3. Panelist Agenda for Grade 5 Science
4. ACT Agenda for ELA, Math, and Science for High School
5. Vertical Articulation Agenda
6. Sample Facilitator Agenda



Standard Setting Agenda

July 14-15, 2025

ELA & Math Grades 4, 6, 8

Science Grade 8

Day 1: Monday July 14, 2025

Start Time	Activity
7:15 a.m.	Registration and breakfast
8:00 a.m.	General Session <ul style="list-style-type: none">• Welcome and introductions• Overview of IAR & ISA• Purpose of the standard setting meeting
9:15 a.m.	Report to breakout session rooms
9:30 a.m.	Breakout Session <ul style="list-style-type: none">• Welcome & Introductions• Orientation to materials and website• Security discussion• Experience the Assessment• Discuss Performance Level Descriptors
12:00 p.m.	Lunch
1:00 p.m.	Develop Borderline Descriptors
2:30 p.m.	Break
2:45 p.m.	Standard Setting Training
3:15 p.m.	Practice Judgment Task Discuss Practice Judgment Task
4:00 p.m.	End of Day

Day 2: Tuesday July 15, 2025

Start Time	Activity
7:15 a.m.	Breakfast
8:00 a.m.	Readiness Survey
8:45 a.m.	Round 1 Judgments
10:15 a.m.	Break
10:30 a.m.	Round 1 Feedback Round 2 Judgments
12:15 p.m.	Lunch
1:15 p.m.	Round 2 Feedback Round 3 Judgments
3:00 p.m.	Break
3:15 p.m.	Round 3 Feedback Final Evaluation
4:00 p.m.	End of Day



Standard Setting Agenda

July 16-17, 2025

ELA & Math – Grades 3, 5, & 7

Day 3: Wednesday July 16, 2025

Start Time	Activity
7:15 a.m.	Registration and breakfast
8:00 a.m.	Breakout Session <ul style="list-style-type: none">• Experience the Assessment (2nd assessment)• Discuss Performance Level Descriptors
10:00 a.m.	Develop Borderline Descriptors
12:00 p.m.	Lunch
1:00 p.m.	Readiness Survey
1:15 p.m.	Round 1 Judgments
3:00 p.m.	Break
3:15 p.m.	Round 1 Feedback
4:00 p.m.	End of Day

Day 4: Thursday July 17, 2025

Start Time	Activity
7:15 a.m.	Breakfast
8:00 a.m.	Round 2 Judgments
9:45 a.m.	Break
10:00 a.m.	Round 2 Feedback Round 3 Judgments
12:00 p.m.	Lunch
1:00 p.m.	Round 3 Feedback Final Evaluation
2:00 p.m.	Optional Discussion on Samples to Success Documents
3:00 p.m.	Adjourn



Standard Setting Agenda

July 16-17, 2025

Science Grade 5

Day 1: Wednesday July 16, 2025

Start Time	Activity
7:15 a.m.	Registration and breakfast
8:00 a.m.	General Session (in Grade 5 Break-Out Room) <ul style="list-style-type: none">• Welcome and introductions.• Overview of IAR & ISA• Purpose of the standard setting meeting
9:15 a.m.	Break
9:30 a.m.	Breakout Session <ul style="list-style-type: none">• Welcome & Introductions• Orientation to materials and website• Security discussion• Experience the Assessment• Discuss Performance Level Descriptors
12:00 p.m.	Lunch
1:00 p.m.	Develop Borderline Descriptors
2:30 p.m.	Break
2:45 p.m.	Standard Setting Training
3:15 p.m.	Practice Judgment Task Discuss Practice Judgment Task
4:00 p.m.	End of Day



Day 2: Thursday July 17, 2025

Start Time	Activity
7:15 a.m.	Breakfast
8:00 a.m.	Readiness Survey
8:45 a.m.	Round 1 Judgments
10:15 a.m.	Break
10:30 a.m.	Round 1 Feedback Round 2 Judgments
12:15 p.m.	Lunch
1:15 p.m.	Round 2 Feedback Round 3 Judgments
3:00 p.m.	Break
3:15 p.m.	Round 3 Feedback Final Evaluation
4:00 p.m.	End of Day



Standard Setting Agenda

ACT Grade 11 (Hybrid Angoff/Empirical Standard Setting)
July 15-17, 2025

Day 2: Tuesday, July 15, 2025

Start Time	Activity
7:15 a.m.	<i>Check-in and breakfast</i>
8:00 a.m.	General Session Welcome and Introductions Brief Overview of the ACT Purpose of the Standard Setting Meeting
9:30 a.m.	<i>Report to breakout session rooms</i>
9:40 a.m.	Breakout Session Welcome Introductions Orientation to materials and website Security discussion Experience the Assessment <i>Break after panelists complete this task</i> Discuss Performance Level Descriptors
12:00 p.m.	<i>Lunch</i>
1:00 p.m.	Discuss Borderline Descriptors Standard Setting Training
2:30 p.m.	<i>Break</i>
2:45 p.m.	Practice Judgment Task Discuss Practice Judgment Task
4 p.m.	<i>Adjourn</i>



Standard Setting Agenda

ACT Grade 11 (Hybrid Angoff/Empirical Standard Setting)
July 15-17, 2025

Day 3: Wednesday, July 16, 2025

Start Time	Activity
7:15 a.m.	<i>Check-in and breakfast</i>
8:00 a.m.	Review/Debrief Readiness Survey
8:15 a.m.	Round 1 Judgments (Angoff Y/N)
9:45 a.m.	<i>Break</i>
10:00 a.m.	Round 1 Feedback Discussion Overview of Empirical Methodology ACT Benchmarks Data and Probabilities of Success
12:00 p.m.	<i>Lunch</i>
1:00 p.m.	Continue Discussion
2:30 p.m.	<i>Break</i>
2:45 p.m.	Continue Discussion Round 2 Judgments (Angoff Y/N) Round 2 Results and Feedback/Discussion
4:00 p.m.	<i>Adjourn</i>



Standard Setting Agenda

ACT Grade 11 (Hybrid Angoff/Empirical Standard Setting)
July 15-17, 2025

Day 4: Thursday, July 17, 2025

Start Time	Activity
7:15 a.m.	<i>Check-in and Breakfast</i>
8:00 a.m.	Review/Debrief Introduce Data Books Discussion
10:15 a.m.	<i>Break</i>
10:30 a.m.	Discussion Training and Round 3 Judgments (Empirical)
12:00 p.m.	<i>Lunch</i>
1:00 p.m.	Round 3 Results Discussion Coherence across Subjects and Methodologies Final (Round 4) Judgments (Empirical)
2:30 p.m.	<i>Break and Evaluation Form</i>
2:45 p.m.	Present Final Recommendations & Concluding Remarks
3:00 p.m.	<i>Adjourn</i>



Standard Setting Agenda

July 18, 2025

Illinois Assessment for Readiness (IAR) - ELA & Math: Grades 3-8

Illinois Science Assessment (ISA) - Grades 5 & 8

ACT- Grades 9-11

Day 5: Friday July 18, 2025

Start Time	Activity
7:15 a.m.	Breakfast
8:00 a.m.	Break Out Room Discussions: ELA, Math, Science <ul style="list-style-type: none">• Introductions• Purpose of Vertical Articulation• Cross Grade PLD Review
9:00 a.m.	Cross Grade Impact Data Review
10:00 a.m.	Break
10:15 a.m.	Recommended Changes to Round 3 Results
11:30 a.m.	Lunch
12:30 p.m.	General Session: Cross-Content Area Impact Data Reasonableness Discussion
1:15 p.m.	Evaluation & Wrap Up
1:30 p.m.	Adjourn

Illinois Standard Setting Facilitator Agenda

July 14–15, 2025

ELA & Math Grades 3–8, Science Grade 8

This agenda is designed for facilitators running one of the seven concurrent standard setting meetings for the Illinois Assessment of Readiness (IAR) and Illinois Science Assessment (ISA). It ensures consistency across meetings for ELA (Grades 3–8), Math (Grades 3–8), and Science (Grade 8). Facilitators should follow the agenda closely, using the guidance provided to lead discussions, manage materials, and maintain security protocols. Please reach out to Tracy Gardner, Ph.D., Senior Measurement Advisor and Standard Setting Lead, if you have any questions about the agenda or meeting. Yong Luo, Ph.D., will also be available on site for technical questions related to the data.

Day 1: Monday, July 14, 2025

7:15 a.m. – Registration and Breakfast

- Pearson Program Team Actions:
 - Ensure registration tables are set up and staffed.
 - Verify that all panelists sign in and receive name badges and folders containing key materials (e.g., Performance Level Descriptors, Item Maps, Item Notes sheet).
 - Confirm breakfast is available in the main meeting area.

8:00 a.m. – General Session (Slides 1–50)

- Activities:
 - Welcome and introductions by Tracy Gardner and ISBE representatives.
 - Overview of IAR and ISA assessments.
 - Purpose of the standard setting meeting.
 - Security and confidentiality protocols.
 - Workshop policies and engagement principles.

- Facilitator Actions:
 - Support Tracy Gardner and ISBE by ensuring panelists are seated.
 - Note key points from the General Session for reference in breakout sessions:
 - Purpose: Align proficiency expectations across ELA, Math, and Science for Grades 3-11, establishing cut scores for performance levels (Below Proficient, Approaching Proficient, Proficient, Above Proficient).
 - ISBE Team (Slide 4): Rae Clementz (Executive Director), Angela Foxhall (Director), Heather Colwell (Math Content Lead), Sarah Leonard (ELA Content Lead), John Hicks (Science Content Lead).
 - Pearson Team (Slide 6): Tracy Gardner (Senior Measurement Advisor, Standard Setting Lead), Yong Luo (Lead Research Scientist), Gulsah Gurkan-Kilicoglu (Primary Research Scientist), Mary Allen (Program Manager), Angie Wisner (Project Manager, Psychometrics), Lise Scheuermann (Math Content Lead), Tiffany Hartmann (ELA Content Lead), Christopher Altermatt (Science Content Lead).
 - Assessments (Slide 11): IAR covers ELA and Math (Grades 3-8), ISA covers Science (Grades 5 & 8), and ACT Suite covers Grades 9-11. The ACT meeting begins on Tuesday.
 - Workshop Policies (Slides 46-49): Emphasize active participation, curiosity, collaboration, and prohibiting mobile device use, removing secure materials, or discussing materials outside the meeting.
 - Engagement Principles (Slide 48): Build trust, value perspectives, engage in debate, commit to purpose, support each other, and adapt.
 - Security (Slides 5, 49): Reinforce confidentiality agreements and restrictions on discussing or removing materials.
 - Standard Setting Overview (Slides 34-40): Uses a standard process with three rounds of judgments, focusing on borderline performance; single-point questions answer (Yes/No); multi-point items consider the most common response for borderline performance.

9:15 a.m. – Report to Breakout Session Rooms

- Facilitator Actions:
 - Help direct panelists to their assigned breakout rooms for their specific grade and subject (e.g., Grade 5 Math, Grade 8 ELA, Grade 8 Science etc.).
 - Ensure rooms are ready to go with panelist logins already logged into standard setting website.

9:30 a.m. – Breakout Session

- Activities:
 - Welcome & Introductions
 - Orientation to materials and website
 - Security discussion
 - Experience the Assessment
 - Discuss Performance Level Descriptors (PLDs)
- Facilitator Actions:
 - Welcome & Introductions:
 - Ask panelists to introduce themselves, sharing their name, region of Illinois, position, years of experience, and a fun question: "When you were in elementary school, what did you want to be when you grew up?"
 - Foster collaboration by emphasizing diverse perspectives.
 - Orientation to Materials and Website:
 - Distribute folders with hard copies of materials (e.g., PLDs, Item Maps) and explain their use.
 - Instruct panelists to use laptops only for meeting-related work and to access the standard setting website for reviewing test items, submitting judgments, and responding to surveys.
 - Demonstrate website navigation, showing how to access the Test Map and Experience the Assessment Items.
 - Remind panelists to return folders at the end of each day.
 - Security Discussion:

- Reinforce confidentiality rules, ensuring panelists understand what materials cannot be removed or discussed outside the session.
- Experience the Assessment:
 - Explain the purpose: Understand the test’s structure, items, administration materials, and scoring from a student’s perspective.
 - Direct panelists to the website to open the "Test Map" and "Experience the Assessment Items."
 - Allocate 60 minutes for panelists to review items, rubrics, and sample responses, taking notes on the Item Notes sheet.
 - Instruct panelists to focus on item complexity, required skills/knowledge, and rubric expectations, not their own performance.
 - Panelists will take a break as they finish their independent review of the assessment.
- Discuss Performance Level Descriptors (PLDs):
 - Define PLDs: Descriptions of knowledge and skills students at various performance levels are expected to display, showing increasing content difficulty and cognitive level.
 - Direct panelists to open PLDs on the website and review them for 10-15 minutes.
 - Facilitate a group discussion using these questions:
 - How do expectations increase from lower to higher performance levels?
 - Are there statements within a performance level that differentiate high-end vs. low-end performance?
 - How different is performance at the bottom of one level (e.g., lowest Proficient) compared to the top of the next lower level (e.g., highest Approaching Proficient)?
 - Notify panelists when one minute remains for PLD review.

12:00 p.m. – Lunch

1:00 p.m. – Discuss Borderline Descriptions

- Facilitator Actions:
 - 1:00 p.m. – Discuss Borderline Descriptors: This section will now focus on a discussion about the nuances that delineate the key distinguishing characteristics of borderline performance at each performance level. Panelists will work in small groups to discuss the key distinguishing characteristics of “just barely” performance compared to performance at other points on the continuum of performance at that level. They will start with Proficient, then Approaching Proficient, and finally Above Proficient. The discussion will cover: What are the key distinguishing characteristics of “just barely” performance compared to performance at other points on the continuum of performance at that level?

2:30 p.m. – Break

- Facilitator Actions:
 - Suggest panelists return materials to their folders before the break.

2:45 p.m. – Standard Setting Training

- Facilitator Actions:
 - Provide an overview of the Modified Angoff process, emphasizing recommending cut scores for each performance level.
 - Explain the judgment process:
 - Single-Point Items: Panelists determine if most students at the lower end (e.g., the borderline) of each performance level would answer correctly (Yes/No).
 - Multi-Point CR Items: Panelists assign the most likely score point for students at the lower end of each performance level, based on rubrics and sample responses.

- Highlight the importance of reviewing item maps and PLDs (Performance Level Descriptors) during the judgment process. Panelists should be reminded to consider a variety of students who perform at the lower end of the performance during judgments.

3:15 p.m. – Practice Judgment Task and Discussion

- Facilitator Actions:
 - Direct panelists to the website to access the practice judgment form and review item sets.
 - Instruct panelists to make judgments for single-point and multi-point items, considering PLDs and a variety of students who perform at the lower end of the performance for Approaching Proficient, Proficient, and Above Proficient.
 - Facilitate group discussion after completion:
 - Review the most popular judgments for each performance level for the first and second item sets.
 - Discuss agreement or spread in judgments, asking panelists to share rationales.
 - For CR items, discuss judgments for each component, focusing on score point patterns and reasons for selections.
 - Check if judgment patterns align with expected score progressions across performance levels.
 - Address questions about the judgment process.

4:00 p.m. – End of Day

- Facilitator Actions:
 - Administer Process Evaluation #1 survey to capture feedback on training and support.
 - Verify that all panelists submit the survey and notify you upon completion.
 - Instruct panelists to log out of the website, close all browsers/tabs, and return all documents to their folders.
 - Collect folders and ensure they are checked in.

- Remind panelists of the 7:15 a.m. breakfast and 8:00 a.m. start time for Day 2.

Day 2: Tuesday, July 15, 2025

7:15 a.m. – Breakfast

- Facilitator Actions:
 - Have breakfast and ensure all computers are ready for panelist login upon start of day.

8:00 a.m. – Readiness Survey

- Facilitator Actions:
 - Direct panelists to complete the Judgment Readiness Quiz on the website.
 - Instruct panelists to answer the two questions and submit the quiz.
 - If any panelist answers "No", address their concerns individually to ensure readiness.

8:45 a.m. – Round 1 Judgments

- Facilitator Actions:
 - Open the PLDs from the website for reference.
 - Direct panelists to the Round 1 Judgment Survey on the website.
 - Instruct panelists to work individually, answering for each item:
 - Single-Point Items: Would most students at the lower end of the specific proficiency level get this item correct?
 - Multi-Point CR Items: Which score point most likely represents the most common response for students at the lower end of this proficiency level?
 - Monitor progress and ensure panelists submit their judgments.

10:15 a.m. – Break

- Facilitator Actions:
 - If you leave during the break, make sure the room is secure.

10:30 a.m. – Round 1 Feedback and Round 2 Judgments

- Facilitator Actions:
 - Round 1 Feedback:
 - Share individual feedback (judgments, cut scores) and committee-level feedback (cut score statistics, panelist agreement).
 - Present item performance data: conditional p-values for single-point items and conditional item means for multi-point items.
 - Facilitate small-group discussions using reflection questions:
 - Compare individual cut scores with committee statistics.
 - Review flagged items with high disagreement, using p-values/item means.
 - Identify reasons for higher/lower expectations among panelists.
 - Explore variations in borderline performance perceptions.
 - Emphasize that consensus is not required, but aim for a common understanding of the performance level for a variety of students at the lower end of the category.
 - Round 2 Judgments:
 - Administer the Judgment Readiness Quiz for Round 2.
 - Direct panelists to the Round 2 Judgment Survey and repeat the judgment process, incorporating Round 1 feedback.

12:15 p.m. – Lunch

- Facilitator Actions:
 - Make sure the room is locked and secure before leaving the area.

1:15 p.m. – Round 2 Feedback and Round 3 Judgments

- Facilitator Actions:
 - Round 2 Feedback:
 - Share individual and committee-level feedback, including cut score statistics and agreement data.
 - Present impact data (percentage of students per performance level) and facilitate discussion using questions:
 - What are your impressions of the data?
 - Does the percentage at Proficient or Above align with expectations?
 - What does the Below Proficient percentage suggest about student readiness?
 - What might the Approaching Proficient percentage indicate?
 - Conduct a full-committee discussion on trends, followed by small-group discussions that go deeper into the data.
 - Finish with a full committee discussion with individuals reporting back on small-group discussions.
 - Round 3 Judgments:
 - Administer the Judgment Readiness Quiz for Round 3.
 - Direct panelists to the Round 3 Judgment Survey and repeat the judgment process, incorporating Round 2 feedback.

3:00 p.m. – Break

3:15 p.m. – Round 3 Feedback and Final Evaluation

- Facilitator Actions:
 - Round 3 Feedback:
 - Review committee-level statistics, panelist agreement graphs, and impact data.
 - Address any questions or concerns from panelists.
 - Final Evaluation:

- Administer Process Evaluation #2 survey to capture feedback on the standard setting process and final cut score recommendations.
- Verify survey submissions and address any final questions.

4:00 p.m. – End of Day

- Facilitator Actions:
 - Explain next steps: Recommendations will be reviewed by ISBE for final approval and used for reporting.
 - Thank panelists for their dedication and contributions.
 - Instruct panelists to log out, close browsers, and return all documents to their folders.
 - Collect folders and ensure all materials are accounted for.

Note: Only one facilitator agenda is included here, but every committee had a similar agenda.

Appendix B: Score Profile Example

ELA

Ability	max_score	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
		650-682	683-703	704-718	719-730	731-740	741-749	750-760	761-770	771-786	787-850
	2	↓	→	→	→	→	→	↑	↑	↑	↑
	2	↓	→	→	→	→	→	↑	↑	↑	↑
	2	↓	↓	↓	↓	↓	→	→	→	→	→
	2	↓	→	→	→	→	→	→	→	→	↑
	2	↓	↓	↓	→	→	→	→	→	→	→
	4	↓	↓	↓	→	→	→	→	↑	↑	↑
	3	↓	→	→	→	→	→	→	→	→	↑
	2	↓	→	→	→	→	→	→	→	→	↑
	2	↓	→	→	→	→	→	→	→	→	→
	2	↓	↓	→	→	→	→	→	↑	↑	↑
	2	↓	↓	→	→	→	→	→	→	→	→
	2	↓	→	→	→	→	→	→	→	→	→
	2	↓	→	→	→	→	→	→	→	→	→
	2	↓	→	→	→	→	→	→	→	→	→
	2	↓	↓	↓	→	→	→	→	→	→	→
	2	↓	→	→	→	→	→	→	→	→	→
	2	↓	↓	↓	↓	↓	→	→	→	→	→
	4	↓	→	→	→	→	→	→	↑	↑	↑
	3	↓	↓	↓	↓	↓	→	→	→	↑	↑
Writing Prompts (PCR)											
	4	●	●	●	●	●	●	●	●	●	●
	3	●	●	●	●	●	●	●	●	●	●
	4	●	●	●	●	●	●	●	●	●	●
	3	●	●	●	●	●	●	●	●	●	●

Display each icon according to these rules:

Icon: ↑ when value is \geq 1.5
 Icon: → when $<$ 1.5 and \geq 0.5
 Icon: ↓ when $<$ 0.5

Display each icon according to these rules:

Icon: ● when value is \geq 2.5
 Icon: ● when $<$ 2.5 and \geq 1.5
 Icon: ● when $<$ 1.5 and \geq 0.5
 Icon: ● when $<$ 0.5

Math

Ability	Max	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
		650-680	681-699	700-709	710-716	717-723	724-731	732-741	742-756	757-773	774-850
	1	→	↑	↑	↑	↑	↑	↑	↑	↑	↑
	1	→	↑	↑	↑	↑	↑	↑	↑	↑	↑
	6	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	1	↓	↓	↓	→	↑	↑	↑	↑	↑	↑
	1	↓	→	↑	↑	↑	↑	↑	↑	↑	↑
	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
	1	↓	↓	→	→	→	↑	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	→	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	1	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	2	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
	3	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	1	↓	↓	↓	↓	↓	↓	→	↑	↑	↑
	2	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
	1	↓	↓	↓	↓	→	→	↑	↑	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	1	↓	↓	↓	↓	↓	→	↑	↑	↑	↑
	2	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑
	1	↓	↓	↓	→	↑	↑	↑	↑	↑	↑
	4	↓	↓	↓	↓	↓	↓	↓	→	→	↑
	1	↓	↓	↓	↓	↓	→	→	→	→	↑
	1	↓	↓	↓	↓	↓	↑	↑	↑	↑	↑
	1	↓	↓	↓	↓	→	↑	↑	↑	↑	↑
	3	↓	↓	↓	↓	↓	↓	→	↑	↑	↑
	1	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑
	1	↓	↓	↓	↓	↓	↑	↑	↑	↑	↑
	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑

Max	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
6	●	●	●	●	●	●	●	●	●	●
3	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●
3	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●
3	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●
4	●	●	●	●	●	●	●	●	●	●
3	●	●	●	●	●	●	●	●	●	●
2	●	●	●	●	●	●	●	●	●	●

Display each icon according to these rules:

Icon: when value is

Icon: when < 0.55 and

Icon: when < 0.45

Display each icon according to these rules:

Icon: when value is

Icon: when < 2.5 and

Icon: when < 1.5 and

Icon: when < 0.5

Science

G5	UIN	Points	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Item	Unit	Scaled Score Range	700-763	764-773	774-785	786-794	795-804	805-812	813-821	822-831	832-843	833-900
51	3		1	↓	→	↑	↑	↑	↑	↑	↑	↑
52	3		1	↓	↑	↑	↑	↑	↑	↑	↑	↑
53	3		1	↓	↓	→	↑	↑	↑	↑	↑	↑
54	3		1	↓	→	↑	↑	↑	↑	↑	↑	↑
55	3		1	↓	↓	↓	↓	↓	↓	↓	→	↑
56	3		1	↓	↓	↓	↓	→	↑	↑	↑	↑
57	3		1	↓	↓	↓	→	→	↑	↑	↑	↑
58	3		1	↓	↓	→	↑	↑	↑	↑	↑	↑
59	3		1	↓	↓	↓	↓	→	↑	↑	↑	↑
60	3		1	↓	↓	↓	↓	↓	→	↑	↑	↑
61	3		1	↓	↓	↑	↑	↑	↑	↑	↑	↑
62	3		1	↓	↓	↓	→	→	→	↑	↑	↑
63	3		1	↓	↓	↓	↓	→	↑	↑	↑	↑
64	3		1	↓	↓	↓	↓	↓	↓	↓	→	↑
65	3		1	↓	↓	→	↑	↑	↑	↑	↑	↑
66	3		1	↓	↓	↓	↑	↑	↑	↑	↑	↑
67	3		1	↓	→	↑	↑	↑	↑	↑	↑	↑
68	3		1	↓	↓	↓	→	↑	↑	↑	↑	↑
69	3		1	↓	↓	↓	↑	↑	↑	↑	↑	↑
70	3		1	↓	↓	↓	→	↑	↑	↑	↑	↑
71	3		1	↓	↓	↓	→	↑	↑	↑	↑	↑
72	3		1	↓	↓	→	↑	↑	↑	↑	↑	↑
73	3		1	↓	↓	↓	↓	→	↑	↑	↑	↑
74	3		1	↓	↓	↓	↓	→	↑	↑	↑	↑
75	3		3	↓	↓	↓	↓	→	→	↑	↑	↑

Item	Unit	Points	700-763	764-773	774-785	786-794	795-804	805-812	813-821	822-831	832-843	833-900
25	1	3	●	●	●	●	●	●	●	●	●	●
50	2	3	●	●	●	●	●	●	●	●	●	●
75	3	3	●	●	●	●	●	●	●	●	●	●

Icon: when value is

Icon: when < 55 and

Icon: when < 45

Display each icon according to these rules:

Icon: when value is

Icon: when < 2.5 and

Icon: when < 1.5 and

Icon: when < 0.5

Appendix C: ACT Data Books

ELA

State Test Performance for Previous Years/Grades

Math Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	5	4	6	7
Met	28	26	27	27
Approached	46	47	52	52
Partially Met	71	71	48*	48*
Did Not Meet*	29	29	--	--

Science Achievement Level	Percent At or Above			
	ISA Grade 8		ISA Grade 11	
	2024	2023	2024	2023
Exemplary	20	19	15	16
Proficient	55	52	52	52
Developing	84	80	79	77
Emerging*	17	20	21	23

ELA Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	9	7	11	11
Met	50	41	32	32
Approached	72	66	67	69
Partially Met	86	85	34*	31*
Did Not Meet*	14	15	--	--

* Did Not Meet and Emerging values are Percent At (not Percent At or Above)

* SAT Partially Met values are Percent At (not Percent At or Above)

https://www.illinoisreportcard.com/State.aspx?source=trends&source2=isa_details&Stateid=IL

Math

State Test Performance for Previous Years/Grades

Math Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	5	4	6	7
Met	28	26	27	27
Approached	46	47	52	52
Partially Met	71	71	48*	48*
Did Not Meet*	29	29	--	--

Science Achievement Level	Percent At or Above			
	ISA Grade 8		ISA Grade 11	
	2024	2023	2024	2023
Exemplary	20	19	15	16
Proficient	55	52	52	52
Developing	84	80	79	77
Emerging*	17	20	21	23

ELA Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	9	7	11	11
Met	50	41	32	32
Approached	72	66	67	69
Partially Met	86	85	34*	31*
Did Not Meet*	14	15	--	--

* Did Not Meet and Emerging values are Percent At (not Percent At or Above)

* SAT Partially Met values are Percent At (not Percent At or Above)

<https://www.illinoisreportcard.com/State.aspx?source=trends&source2=isa.details&Stateid=IL>

Science

State Test Performance for Previous Years/Grades

Math Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	5	4	6	7
Met	28	26	27	27
Approached	46	47	52	52
Partially Met	71	71	48*	48*
Did Not Meet*	29	29	--	--

Science Achievement Level	Percent At or Above			
	ISA Grade 8		ISA Grade 11	
	2024	2023	2024	2023
Exemplary	20	19	15	16
Proficient	55	52	52	52
Developing	84	80	79	77
Emerging*	17	20	21	23

ELA Achievement Level	Percent At or Above			
	IAR Grade 8		SAT Grade 11	
	2024	2023	2024	2023
Exceeded	9	7	11	11
Met	50	41	32	32
Approached	72	66	67	69
Partially Met	86	85	34*	31*
Did Not Meet*	14	15	--	--

* Did Not Meet and Emerging values are Percent At (not Percent At or Above)

* SAT Partially Met values are Percent At (not Percent At or Above)

<https://www.illinoisreportcard.com/State.aspx?source=trends&source2=isa.details&Stateid=IL>

Handouts for Round 1 ACT Feedback Probabilities

Probability of Success and Percentage of Students At or Above Each ACT Score
ACT Subject: Mathematics
College Course: Algebra

ACT Math Score	Probability of Success			Percentage At/Above
	A prob	B or higher prob	C or higher prob	IL Grade 11 ACT 2025
36	0.89	0.94	0.96	1
35	0.86	0.93	0.95	3
34	0.83	0.92	0.95	4
33	0.78	0.90	0.94	5
32	0.74	0.88	0.92	5
31	0.70	0.86	0.91	6
30	0.64	0.83	0.90	7
29	0.58	0.80	0.89	8
28	0.51	0.77	0.87	10
27	0.45	0.73	0.85	12
26	0.39	0.69	0.83	15
25	0.34	0.64	0.80	19
24	0.29	0.59	0.78	23
23	0.23	0.55	0.75	26
22	0.20	0.51	0.73	29
21	0.16	0.46	0.70	32
20	0.13	0.40	0.66	35
19	0.11	0.35	0.63	39
18	0.09	0.30	0.60	45
17	0.07	0.26	0.56	52
16	0.05	0.22	0.51	63
15	0.04	0.19	0.46	75
14	0.03	0.16	0.43	84
13	0.02	0.13	0.39	91
12	0.02	0.11	0.35	95
N-count	70,461			139,151

P(Success) and Impact-Math

Appendix D: Evaluation 1

Summary of Evaluation 1 for Grades 3–8

The Evaluation 1 surveys for the Standard Setting Meeting across grades 3-8 for ELA and Math, and grades 5 and 8 for Science, aimed to assess the usefulness of training activities, the effectiveness of activities in assisting recommendations, and the adequacy of time for training and practice components. Below is a summary of the findings, followed by a table of average scores and a list of open-ended responses.

Key Findings:

- Usefulness of Training Activities:** Most training activities were rated as "Useful" to "Very Useful" across all grades. The highest ratings were often for experiencing the actual assessment and practice activities, indicating that hands-on experience was particularly valued.
- Effectiveness in Assisting Recommendations:** The standard setting training, PLD discussions, and borderline description discussions were generally rated as "Very Useful" in assisting participants to make recommendations. This suggests that these components were effective in preparing participants for their tasks.
- Adequacy of Time:** The time allocated for training and practice components was generally rated as "Adequate" to "More Than Adequate." However, there were some variations, with a few participants suggesting that more time could be beneficial, particularly for discussions.

Activity	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Pre-work activities	3.5	3.6	3.4	3.5	3.7	3.6
Overview of the Illinois assessments	3.5	3.4	3.5	3.6	3.7	3.6
Introduction to the standard setting process	3.6	3.6	3.6	3.7	3.7	3.7
Experiencing the actual assessment	3.8	3.7	3.7	3.8	3.9	3.8
Discussion of the scoring of items	3.7	3.5	3.7	3.8	3.8	3.8
Discussion of the PLDs	3.7	3.5	3.7	3.8	3.8	3.8
Overview of the standard setting procedure	3.6	3.5	3.6	3.8	3.8	3.8
Discussion of the borderline descriptions	3.6	3.5	3.6	3.7	3.8	3.7
Practice activity	3.7	3.6	3.7	3.8	3.9	3.8

ELA	Activity	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Pre-work activities	3.4	3.5	3.3	3.4	3.6	3.5
	Overview of the Illinois assessments	3.4	3.3	3.4	3.5	3.6	3.5
	Introduction to the standard setting process	3.5	3.5	3.5	3.6	3.6	3.6

	Experiencing the actual assessment	3.7	3.6	3.6	3.7	3.8	3.7
	Discussion of the scoring of items	3.6	3.4	3.6	3.7	3.7	3.7
	Discussion of the PLDs	3.6	3.4	3.6	3.7	3.7	3.7
	Overview of the standard setting procedure	3.5	3.4	3.5	3.7	3.7	3.7
	Discussion of the borderline descriptions	3.5	3.4	3.5	3.6	3.7	3.6
	Practice activity	3.6	3.5	3.6	3.7	3.8	3.7

Math							
	Activity	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Pre-work activities	3.6	3.7	3.5	3.6	3.8	3.7
	Overview of the Illinois assessments	3.6	3.5	3.6	3.7	3.8	3.7
	Introduction to the standard setting process	3.7	3.7	3.7	3.8	3.8	3.8
	Experiencing the actual assessment	3.8	3.8	3.8	3.9	3.9	3.9
	Discussion of the scoring of items	3.8	3.6	3.8	3.9	3.9	3.9
	Discussion of the PLDs	3.8	3.6	3.8	3.9	3.9	3.9
	Overview of the standard setting procedure	3.7	3.6	3.7	3.9	3.9	3.9
	Discussion of the borderline descriptions	3.7	3.6	3.7	3.8	3.9	3.8
	Practice activity	3.8	3.7	3.8	3.9	3.9	3.9

SC			
	Activity	Grade 5	Grade 8
	Pre-work activities	3.3	3.5
	Overview of the Illinois assessments	3.4	3.5
	Introduction to the standard setting process	3.5	3.6
	Experiencing the actual assessment	3.6	3.7
	Discussion of the scoring of items	3.6	3.7
	Discussion of the PLDs	3.6	3.7
	Overview of the standard setting procedure	3.5	3.7
	Discussion of the borderline descriptions	3.5	3.6
	Practice activity	3.6	3.7

Open-Ended Responses:

1. **More Time Needed:** Several participants suggested that more time could be allocated to discussions and practice activities to ensure thorough understanding and engagement.
2. **Appreciation for Hands-On Activities:** Many participants appreciated the opportunity to engage directly with the assessments and practice activities, noting that these were particularly helpful in understanding the standard setting process.
3. **Clarity and Structure:** Some participants mentioned the need for clearer instructions and more structured sessions to enhance the effectiveness of the training.
4. **Positive Feedback on Facilitators:** Participants frequently praised the facilitators for their knowledge and ability to guide discussions effectively.
5. **Suggestions for Improvement:** A few participants suggested incorporating more real-world examples and case studies to make the training more relatable and impactful.

Overall, the Evaluation 1 surveys indicate a positive reception of the standard setting process, with most participants finding the training activities useful and the time allocated generally adequate. The feedback highlights the importance of hands-on activities and suggests areas for potential improvement in future sessions.

Appendix E: Evaluation 2

Summary of Evaluation 2 for Grades 3-8

The Evaluation 2 surveys for the Standard Setting Meeting across grades 3-8 for ELA and Math, and grades 5 and 8 for Science, focused on the success of the judgment process, the usefulness of feedback data, confidence in cut score recommendations, and the adequacy of meeting facilities and interactions. Below is a summary of the findings, followed by tables of average scores by content area and a list of open-ended responses.

Key Findings:

- **Success of Judgment Process:** The judgment rounds and feedback discussions were generally rated as "Successful" to "Very Successful" across all grades and subjects. Participants appreciated the structured process and the opportunity to engage in meaningful discussions.
- **Usefulness of Feedback Data:** Feedback data, including individual item judgments, cut scores, and committee statistics, were rated as "Useful" to "Very Useful." Participants found the data helpful in making informed decisions.
- **Confidence in Cut Scores:** Confidence in the final cut score recommendations was generally high, with most participants feeling "Confident" to "Very Confident" about the recommendations representing student achievement.
- **Adequacy of Facilities and Tools:** The meeting facilities and tools were rated as "Adequate" to "More Than Adequate," though some participants noted issues with room temperature and suggested improvements in materials and resources.

Average Scores by Content Area and Grade

ELA							
	Activity	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Judgment Rounds 1-3	3.8	3.7	3.9	3.8	3.5	4.0
	Feedback data after each round	3.9	3.8	3.8	3.9	3.7	4.0
	Discussion after each round	3.9	3.8	3.9	3.8	3.5	3.9
	Confidence in cut scores	3.7	3.6	3.7	3.7	3.5	3.8
	Facilities and tools	3.8	3.7	3.9	3.8	3.5	3.9

Math							
	Activity	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Judgment Rounds 1-3	3.7	3.6	3.8	3.7	3.5	3.9
	Feedback data after each round	3.8	3.7	3.8	3.8	3.6	3.9
	Discussion after each round	3.8	3.7	3.8	3.7	3.5	3.8
	Confidence in cut scores	3.6	3.5	3.7	3.6	3.4	3.7
	Facilities and tools	3.7	3.6	3.8	3.7	3.4	3.8

Science			
	Activity	Grade 5	Grade 8
	Judgment Rounds 1-3	2.9	3.8
	Feedback data after each round	3.0	3.9
	Discussion after each round	2.8	3.9
	Confidence in cut scores	3.0	3.7
	Facilities and tools	3.4	3.8

Narrative Summary

Overall, the Evaluation 2 surveys indicate a successful standard setting process, with participants finding the judgment rounds and feedback data useful and expressing confidence in the final cut score recommendations. The facilities and tools were generally adequate, though some participants suggested improvements in room conditions and materials. The feedback highlights the effectiveness of the process and suggests areas for enhancement in future sessions.

Open-Ended Responses

1. **More Time Needed:** Several participants suggested that more time could be allocated to discussions and feedback rounds to ensure thorough understanding and engagement.
2. **Appreciation for Facilitators:** Many participants praised the facilitators for their knowledge and ability to guide discussions effectively.
3. **Suggestions for Improvement:** Participants suggested incorporating more practice sessions and ensuring consistent information across sessions to enhance the process.
4. **Positive Experience:** Many participants expressed appreciation for the opportunity to participate and found the experience valuable and informative.

This comprehensive summary and breakdown by content area provide a clear picture of the evaluation results and participant feedback, highlighting the strengths and areas for improvement in the standard setting process.

Appendix F: ACT Evaluations

ACT Evaluation 1

Evaluation 1 was administered at the end of the first day of the ACT standard setting (Tuesday), after participants completed the practice activity but had not engaged in any of the Rounds of judgments. Overall, panelists found the standard setting training activities to be useful, and amount of time spent on each activity to be adequate.

How useful were the following standard setting training activities?	Very Useful	Useful	Somewhat Useful	Not Useful	Useful or Very Useful	Mean
Overview of the assessments	65%	28%	7%	0%	93%	3.6
Introduction to the standard setting process	85%	15%	0%	0%	100%	3.8
Experiencing the assessment	80%	15%	4%	0%	96%	3.8
Discussion of the PLDs	83%	17%	0%	0%	100%	3.8
Discussion of the borderline descriptions	80%	15%	4%	0%	96%	3.8
Practice Activity	93%	7%	0%	0%	100%	3.9

How useful were the following activities in assisting you to make your recommendations?	Very Useful	Useful	Somewhat Useful	Not Useful	Useful or Very Useful	Mean
Standard setting training	78%	22%	0%	0%	100%	3.8
PLD discussion	87%	13%	0%	0%	100%	3.9
Borderline description discussion	78%	20%	2%	0%	98%	3.8

How adequate was the amount of time spent on the following training and practice components	Very Adequate	Adequate	Somewhat Adequate	Not Adequate	Adequate or Very Adequate	Mean
Standard setting training	67%	33%	0%	0%	100%	3.7
PLD discussion	74%	24%	2%	0%	98%	3.7
Borderline description discussion	59%	35%	7%	0%	93%	3.5
Discussion of practice judgments	57%	39%	4%	0%	96%	3.5

ACT Evaluation 2

Evaluation 2 was administered at the end of the second day of the ACT standard setting (Wednesday), after completing two round of judgments using the Extended Modified Angoff Yes/No approach. This evaluation was intended to capture the process so far, and to ensure that the panelists from the English, Reading, and Writing panels who would be dismissed at the end of the day had the opportunity to provide feedback. Overall, panelists found the process and feedback data to be useful, the elements of the standard

setting to be adequate, and facilities and tools to be adequate. Panelists were less confident in the cut score recommendations so far, but this is to be expected because the evaluation was administered after only two rounds of judgments and the process was incomplete.

How successful were the following components of the standard setting judgment process?	Very Useful	Useful	Somewhat Useful	Not Useful	Useful or Very Useful	Mean
Judgment Rounds 1-2	57%	30%	13%	0%	88%	3.4
Feedback data after each round	71%	23%	5%	0%	95%	3.7
Discussion after each round	68%	27%	4%	2%	95%	3.6

How useful were the following feedback data after each round?	Very Useful	Useful	Somewhat Useful	Not Useful	Useful or Very Useful	Mean
Individual item judgments	66%	32%	2%	0%	98%	3.6
Individual cut scores	68%	30%	2%	0%	98%	3.7
Committee agreement graphs	79%	18%	4%	0%	96%	3.8
Committee cut score statistics	80%	20%	0%	0%	100%	3.8
Impact data	68%	27%	4%	2%	95%	3.6
Student score profiles	71%	20%	7%	2%	91%	3.6
Probabilities of success in college	57%	25%	14%	4%	82%	3.4

How adequate were the following elements of the standard setting meeting?	Very Adequate	Adequate	Somewhat Adequate	Not Adequate	Adequate or Very Adequate	Mean
Amount of time to make judgments	70%	27%	4%	0%	96%	3.7
Presentation of the feedback data	70%	30%	0%	0%	100%	3.7
Number of judgment rounds	54%	38%	9%	0%	91%	3.4

How confident are you that the borderline descriptions your committee discussed based on the PLDs are reasonable for each performance level?	Very Confident	Confident	Somewhat Confident	Not Confident	Confident or Very Confident	Mean
Approaching Proficient	36%	32%	23%	9%	68%	2.9
Proficient	39%	38%	21%	2%	77%	3.1
Above Proficient	46%	34%	20%	0%	80%	3.3

How confident are you that the cut score recommendations from your committee adequately represent student achievement?	Very Confident	Confident	Somewhat Confident	Not Confident	Confident or Very Confident	Mean
Approaching Proficient	25%	21%	43%	11%	46%	2.6
Proficient	30%	43%	25%	2%	73%	3.0
Above Proficient	32%	41%	20%	7%	73%	3.0

How adequate were the following facilities and tools used during the standard setting meeting?	Very Adequate	Adequate	Somewhat Adequate	Not Adequate	Adequate or Very Adequate	Mean
Facilities used for the meeting	80%	18%	2%	0%	98%	3.8
Computers used during the meeting	84%	16%	0%	0%	100%	3.8
Standard setting website	77%	21%	0%	2%	98%	3.7
Online forms for submitting judgments	88%	13%	0%	0%	100%	3.9
Materials provided in the folder	84%	14%	2%	0%	98%	3.8
Workspace in table groups	61%	25%	9%	5%	86%	3.4

How adequately were you able to complete the following activities during the standard setting?	Very Adequate	Adequate	Somewhat Adequate	Not Adequate	Adequate or Very Adequate	Mean
Express your opinions	86%	14%	0%	0%	100%	3.9
Ask questions and clarification about the cut scores	82%	18%	0%	0%	100%	3.8
Ask questions about the standard setting process	86%	14%	0%	0%	100%	3.9
Interact with your fellow panelists	84%	13%	0%	4%	96%	3.8

How adequately were your opinions and judgments treated with respect by the facilitators and other panelists?	Very Adequate	Adequate	Somewhat Adequate	Not Adequate	Adequate or Very Adequate	Mean
Fellow panelists	79%	20%	2%	0%	98%	3.8
Facilitators	95%	5%	0%	0%	100%	3.9

ACT Evaluation 3

Evaluation 3 was administered at the end of the third day of the ACT standard setting (Thursday) to capture final feedback about the hybrid approach and the final cut score recommendations before the vertical articulation. Overall, panelists indicated that they understood the purpose and activities of the standard setting event, they were comfortable using the concepts and data presented, they did not feel pressured when making their cut score determinations, they felt that their input was valued, they found the process to be highly effective and the resulting cut scores to be reasonable and defensible, and that the amount of time allotted to each round was adequate. In open-ended comments, panelists indicated that they appreciated the opportunity to participate in the standard setting, they found the process to be valuable and well organized, they appreciated the facilitators, and a couple of comments specifically highlighted the value of the hybrid approach, finding the combination of both content and empirical data to be particularly valuable when making their cut score determinations.

How well did you understand the following aspects of the standard setting?	Very Well	Well	Somewhat Well	Not Well	Well or Very Well	Mean
The purpose of the meeting	86%	14%	0%	0%	100%	3.9
The instructions on what you were to do each round	92%	8%	0%	0%	100%	3.9
The tasks you were to accomplish each round	94%	6%	0%	0%	100%	3.9
The difference between borderline and typical performance within a performance level	89%	11%	0%	0%	100%	3.9
The median cut scores	89%	11%	0%	0%	100%	3.9
The concept of using performance in first-year credit-bearing college courses to help set cut scores	89%	11%	0%	0%	100%	3.9
Probabilities of success	89%	11%	0%	0%	100%	3.9
The difference between probability of success and percent at or above	86%	14%	0%	0%	100%	3.9

How comfortable were you with the following concepts?	Very Comfortable	Comfortable	Somewhat Comfortable	Not Comfortable	Comfortable or Very Comfortable	Mean
Performance at the borderline of Approaching Proficient	72%	25%	3%	0%	97%	3.7
Performance at the borderline of Proficient	78%	22%	0%	0%	100%	3.8
Performance at the borderline of Above Proficient	69%	25%	6%	0%	94%	3.6
Using impact data to evaluate the reasonableness of the cut scores	78%	22%	0%	0%	100%	3.8

Did you feel pressured during the standard setting event?	Very Pressured	Moderately Pressured	Slightly Pressured	Not at all Pressured	Slightly or Not At All Pressured	Mean
By others in your group to make your cut score recommendations agree with theirs?	8%	0%	6%	86%	92%	1.3
By staff to make your recommendations higher or lower?	6%	0%	0%	94%	94%	1.2
By staff to keep your recommendations the same?	6%	0%	3%	92%	94%	1.2

	Very Valued	Moderately Valued	Slightly Valued	Not at all Valued	Moderately or Very Valued	Mean
To what extent was your input valued and considered by others in your group?	81%	14%	6%	0%	94%	3.8

	Very Confident	Confident	Somewhat Confident	Not Confident	Confident or Very Confident	Mean
How confident are you that the cut score recommendations from your committee adequately represent student achievement?						
Approaching Proficient	56%	36%	8%	0%	92%	3.5
Proficient	69%	22%	8%	0%	92%	3.6
Above Proficient	56%	28%	17%	0%	83%	3.4

	Very Effective	Effective	Somewhat Effective	Not Effective	Effective or Very Effective	Mean
How effective was the standard setting method?						
Rounds 1-2 (item-level judgments)	64%	31%	6%	0%	94%	3.6
Rounds 3-4 (probabilities of success)	78%	22%	0%	0%	100%	3.8

	Very	Moderately	Slightly	Not at all	Moderately or Very	Mean
How defensible do you feel are the cut scores produced by this standard setting process?						
	86%	14%	0%	0%	100%	3.9
How reasonable do you feel are the cut scores produced by this standard setting process?						
	83%	17%	0%	0%	100%	3.8

How did you feel about the amount of time allotted for explanation and discussion each round?	Far too long	Somewhat long	About right	Somewhat short	Far too short	Somewhat to About Right	Mean
Round 1	0%	8%	69%	19%	3%	97%	2.8
Round 2	3%	11%	75%	8%	3%	94%	3.0
Round 3	6%	19%	72%	0%	3%	92%	3.3
Round 4	0%	11%	86%	0%	3%	97%	3.1

Appendix G: Vertical Articulation Survey

The following survey questions were used to collect feedback from participants in the Vertical Articulation meeting for the Illinois Assessment of Readiness (IAR), Illinois Science Assessment (ISA), and ACT Assessments, held on July 18, 2025. The questions are organized into sections to assess the success of meeting components, adequacy of training elements, understanding and support for cut score recommendations, confidence in cut score recommendations, and open-ended feedback. Questions with similar response options are presented in tables for clarity and ease of inclusion in the Standard Setting report.

LEVEL OF SUCCESS OF MEETING COMPONENTS

Participants were asked to evaluate the success of the following meeting components by selecting the option that best reflected their opinion:

Question	Response Options (1-4)
1. Introduction and training to vertical articulation process	Not Successful, Partially Successful, Successful, Very Successful
2. Review of the Performance Level Descriptors (PLDs)	Not Successful, Partially Successful, Successful, Very Successful
3. Review of the cross-grade impact data	Not Successful, Partially Successful, Successful, Very Successful
4. Use of interactive vertical articulation spreadsheet	Not Successful, Partially Successful, Successful, Very Successful
5. Discussion of recommended changes across grades within the content area I served on	Not Successful, Partially Successful, Successful, Very Successful
6. General Session Review and Discussion: Content reasonableness review across content areas	Not Successful, Partially Successful, Successful, Very Successful

Understanding and Support for Cut Score Recommendations

Participants were asked to respond to the following questions regarding their understanding and support for the cut score recommendations:

Question	Response Options
7. I understood the process for recommending changes to the scaled score cuts for the performance levels in our content area discussion.	Yes, No
8. If you responded "No" to the previous question, please elaborate.	[Open response]
9. I support the cut scores that were recommended after the Vertical Articulation Process for the content area that I served on.	Yes, No
10. Please elaborate on your response to the previous question.	[Open response]
11. I support the FINAL cut scores that were recommended after the Cross Content Area Reasonableness Review.	Yes, No
12. Please elaborate on your response to the previous question.	[Open response]

Adequacy of Vertical Articulation Training Elements

Participants were asked to evaluate the adequacy of the following elements of the vertical articulation session:

Question	Response Options (1-4)
1. Amount of time spent reviewing the PLDs	Not Adequate, Somewhat Adequate, Adequate, More than Adequate
2. Amount of time discussing the impact data across grades	Not Adequate, Somewhat Adequate, Adequate, More than Adequate
3. Amount of time working with the interactive spreadsheet	Not Adequate, Somewhat Adequate, Adequate, More than Adequate
4. Amount of time for content reasonableness review	Not Adequate, Somewhat Adequate, Adequate, More than Adequate

CONFIDENCE IN FINAL CUT SCORE RECOMMENDATIONS

Participants were asked to indicate their level of confidence in the final cut score recommendations for each subject and performance level:

Subject	Performance Level	Response Options (1-4)
ELA	Approaching Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Above Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
Mathematics	Approaching Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Above Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
Science	Approaching Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Proficient	Not Confident, Somewhat Confident, Confident, Very Confident
	Above Proficient	Not Confident, Somewhat Confident, Confident, Very Confident

Open-Ended Questions

Participants were asked to provide detailed responses to the following questions:

- Cut Score Suggestions by Grade and Content Area**
 If you disagree with any final scaled score recommendations after the final content reasonableness review, please explain your concerns here. If applicable, include which subject areas, grades, and performance levels you are concerned with. [Open response]
- Additional Feedback**
 Please provide any additional comments or suggestions to improve the vertical articulation process or the content reasonableness review. [Open response]

Summary of Vertical Articulation Evaluation

The Vertical Articulation Evaluation surveys for ELA, Math, and Science focused on assessing the effectiveness of the vertical articulation process, evaluating the success of meeting components, gauging confidence in final cut score recommendations, and identifying strengths and areas for improvement. Below is a summary of the findings for each content area, followed by tables of average scores and a list of open-ended responses.

ELA VERTICAL ARTICULATION

Key Findings:

- **Success of Meeting Components:** The components, including the introduction to the process, PLD reviews, and cross-grade data analysis, were generally rated as "Successful" to "Very Successful." Participants appreciated the structured process and the opportunity for meaningful discussions.
- **Confidence in Cut Scores:** Confidence in the final cut score recommendations was high, with most participants feeling "Confident" to "Proficient."
- **Adequacy of Training:** The time allocated for reviewing PLDs and discussing impact data was generally considered "Adequate" to "More Than Adequate."

ELA	Activity	Average Score
	Introduction to vertical articulation process	3.6
	Review of the PLDs	3.8
	Review of the cross-grade impact data	3.9
	Use of interactive vertical articulation spreadsheet	3.9
	Discussion of recommended changes	3.8
	General Session Review and Discussion	3.7

Open-Ended Responses:

1. **Positive Feedback:** Participants appreciated the thorough process and the opportunity to contribute to meaningful decisions.
2. **Suggestions for Improvement:** Some participants suggested more time for discussions within grade bands and clearer communication of the process to stakeholders.

MATH VERTICAL ARTICULATION

Key Findings:

- **Success of Meeting Components:** The meeting components were rated as "Successful" to "Very Successful," with participants valuing the structured approach and collaborative discussions.
- **Confidence in Cut Scores:** Participants expressed high confidence in the final cut score recommendations, with unanimous support for the decisions made.
- **Adequacy of Training:** The training elements were generally rated as "Adequate" to "More Than Adequate."

Math		
	Activity	Average Score
	Introduction to vertical articulation process	3.8
	Review of the PLDs	3.8
	Review of the cross-grade impact data	3.7
	Use of interactive vertical articulation spreadsheet	3.7
	Discussion of recommended changes	3.7
	General Session Review and Discussion	3.7

Open-Ended Responses:

1. **Positive Feedback:** Participants appreciated the collaborative nature of the process and the focus on aligning cut scores with PLDs and standards.
2. **Suggestions for Improvement:** Some participants suggested involving more teachers to enhance the validity of the process.

SCIENCE VERTICAL ARTICULATION

Key Findings:

- **Success of Meeting Components:** The components were rated as "Successful" to "Very Successful," with participants appreciating the structured process and the focus on student success.
- **Confidence in Cut Scores:** Confidence in the final cut score recommendations was high, with unanimous support for the decisions made.
- **Adequacy of Training:** The training elements were generally rated as "Adequate" to "More Than Adequate."

Science	Activity	Average Score
	Introduction to vertical articulation process	3.7
	Review of the PLDs	3.7
	Review of the cross-grade impact data	3.5
	Use of interactive vertical articulation spreadsheet	3.7
	Discussion of recommended changes	3.8
	General Session Review and Discussion	3.5

Open-Ended Responses:

1. **Positive Feedback:** Participants valued the opportunity to engage in meaningful discussions and appreciated the focus on aligning cut scores with student success.
2. **Suggestions for Improvement:** Some participants suggested more time for discussions and involving a larger sample of teachers to enhance the process.

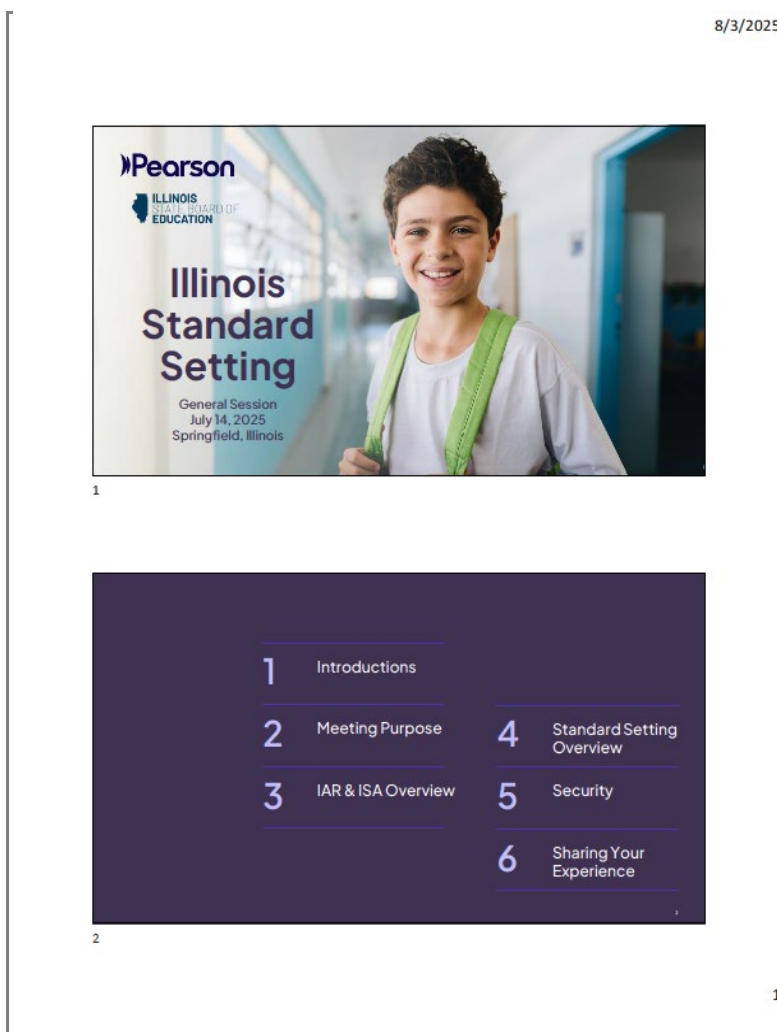
OVERALL SUMMARY

The Vertical Articulation Evaluation surveys indicate a successful process across all content areas, with participants finding the meeting components effective and expressing confidence in the final cut score recommendations. The feedback highlights the importance of structured discussions and suggests areas for potential improvement, such as involving more teachers and providing more time for discussions within grade bands. Participants appreciated the opportunity to contribute to important decisions affecting student assessments and expressed interest in future participation.

Appendix H: Presentations

Click [here](#) to open a document that has the following presentations associated with IAR and ISA.
Note: Depending on the current display settings in your PDF software, the presentations may or may not open in a separate tab.

1. General Session
2. Break Out Presentations for ELA, Math, and Science
3. Vertical Articulation



Click [here](#) to open a document that has the following presentations associated with ACT.

Note: Depending on the current display settings in your PDF software, the presentations may or may not open in a separate tab.

1. General Session
2. Break Out Presentations for ELA, Math, and Science



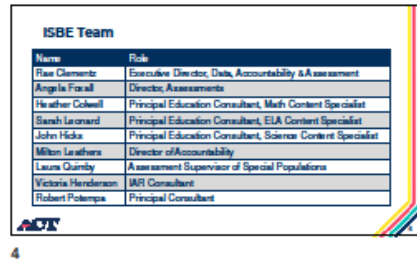
1



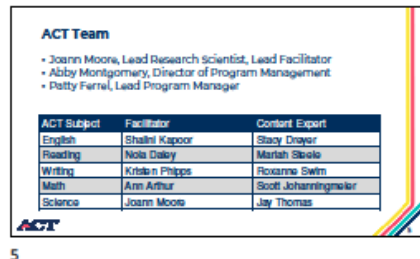
2



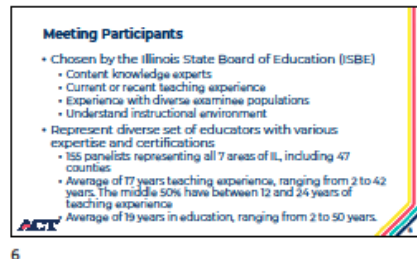
3



4



5



6