



Technical Report

December 2025

Table of Contents

Chapter 1 Statewide System of Standards and Assessments	8
1.1. State Adoption of Academic Content Standards for All Students	8
1.2. Challenging Academic Content Standards	9
1.2.1. Goals of Kansas Standards.....	9
1.2.2. Process and Timeline	10
1.2.3. Standards Review Committees.....	10
1.3. Required Assessments and Intended Population.....	11
1.4. Policies for Including All Students in Assessments	11
Chapter 2 Assessment System Operations.....	13
2.1. Test Design and Development	13
2.1.1. Test Blueprints.....	13
2.1.2. Test Design.....	14
2.1.3. Operational Test Construction.....	15
2.2. Item Development	16
2.2.1. English Language Arts Passage Selection and Review.....	17
2.2.2. Item Writing.....	21
2.2.3. Item Review.....	21
2.3. Test Administration	29
2.3.1. Test Administration and Security Training.....	30
2.3.2. Test Administration Procedures.....	31
2.4. Monitoring Test Administration	32
2.5. Test Security.....	33
2.5.1. Prevention of Test Irregularities	34
2.5.2. Detection of Test Irregularities.....	35
2.5.3. Remediation of Test Irregularities	35
2.6. Protecting Data Integrity and Privacy.....	36
2.6.1. Test Materials Integrity	36
2.6.2. Test-Related Data Integrity and Privacy.....	36
2.6.3. Personally Identifiable Information Integrity and Privacy.....	37
2.6.4. Accommodations-Related Data Integrity and Privacy	37
Chapter 3 Technical Quality: Validity.....	39
3.1. Validity Evidence Based on Test Content	39

3.2. Validity Evidence Based on Internal Structure.....	41
3.2.1. Dimensionality	41
3.2.2. Item Response Theory and Model Assumptions.....	41
3.2.3. Differential Item Functioning.....	44
3.3. Validity Evidence Based on Relations to Other Variables.....	45
3.3.1. Relationships Among KAP Subjects	45
3.3.2. Relationships Between KAP Assessment and National Assessment of Educational Progress	46
3.3.3. Relationships Between KAP Assessment and KELPA.....	48
3.4. Validity Evidence Based on Consequences of Testing	50
Chapter 4 Technical Quality: Other.....	51
4.1. Reliability.....	51
4.1.1. Test Reliability	52
4.1.2. Test Information	55
4.1.3. Classification Consistency and Accuracy.....	60
4.1.4. Subscore Reliability	62
4.2. Accessibility and Fairness	65
4.2.1. Accessibility	65
4.2.2. Fairness	65
4.3. Full Performance Continuum.....	66
4.3.1. Classical Item Statistics.....	66
4.3.2. Item Response Theory Item Statistics	69
4.4. Scoring and Scaling.....	72
4.4.1. Scoring.....	72
4.4.2. Scaling	72
4.4.3. Operational Test Results	75
4.5. Multiple Assessment Forms	91
4.5.1. Cross-Year Linking Design.....	91
4.5.2. Cross-Year Linking Procedure	91
4.6. Multiple Versions of an Assessment.....	92
4.7. Technical Analysis and Ongoing Maintenance	92
Chapter 5 Inclusion of All Students	93
5.1. Procedures for Including Students with Disabilities	93
5.2. Procedures for Including English Learners.....	93

5.3. Accessibility Tools.....	94
5.4. Accommodations.....	95
5.4.1. Selection of Accommodations.....	97
5.4.2. Frequency of Accommodation Use	98
Chapter 6 Academic Achievement Standards and Reporting	99
6.1. State Adoption of Academic Achievement Standards for All Students.....	99
6.2. Achievement Standard Setting.....	99
6.2.1. Standard-Setting Method	99
6.2.2. Procedures	100
6.3. Challenging and Aligned Academic Achievement Standards	116
6.3.1. Development of Grade-Specific Performance Level Descriptors	116
6.3.2. Academic Performance Standards	117
6.4. Reporting.....	118
6.4.1. Student Reports.....	118
6.4.2. School and District Reports.....	119
6.4.3. Reporting Timeline	120
6.4.4. Interpretive Guides	120
References	121
Appendix A Test Blueprints	123
Appendix B Word Count and Lexile Range.....	129
Appendix C Demographic Information for Passage Reviewers	130
Appendix D Demographic Information for External Item Reviewers	136
Appendix E Item Statistics Flagging Criteria	144
Appendix F Demographic Information for Data Reviewers	145
Appendix G Subscore Reliability.....	148
Appendix H 2025 KAP Standard Setting – Opening Session Readiness Evaluation Results	151
Appendix I 2025 KAP Standard Setting – Round 1 Readiness Evaluation Results	153

Table of Tables

Table 2-1. Development Timeline for KAP Assessments	13
Table 2-2. Number of Operational Items on the 2025 KAP Assessment by Subject and Session	15
Table 2-3. External Passage Review Time, Location, and Panelist Counts	19
Table 2-4. ELA External Item Review Time, Location, and Panelist Counts.....	24
Table 2-5. Math External Item Review Time, Location, and Panelist Counts.....	24
Table 2-6. Science External Item Review Time, Location, and Panelist Counts	25
Table 2-7. ELA Data Review Time, Location, and Panelist Counts	27
Table 2-8. Math Data Review Time, Location, and Panelist Counts	28
Table 2-9. Science Data Review Time, Location, and Panelist Counts	28
Table 3-1. Sample Size for Scale-Setting Calibration by Subject and Grade.....	43
Table 3-2. Correlations (C) and Disattenuated Correlations (DC) Among ELA, Mathematics, and Science Scores	46
Table 3-3. Correlations Between KAP ELA Scale Scores and KELPA Domain Scores by Grade	49
Table 3-4. Correlations Between KAP Mathematics Scale Scores and KELPA Domain Scores by Grade	49
Table 3-5. Correlations Between KAP Science Scale Scores and KELPA Domain Scores by Grade.....	50
Table 4-1. Test Reliability Estimate by Subject and Grade.....	52
Table 4-2. Student-Group Reliability Estimates for English Language Arts	53
Table 4-3. Student-Group Reliability Estimates for Mathematics	54
Table 4-4. Student-Group Reliability Estimates for Science	55
Table 4-5. CSEM Across Subjects, Performance Levels, and Grade Levels	60
Table 4-6. Overall Classification Consistency (C) and Accuracy (A).....	61
Table 4-7. Subscores for English Language Arts by Grade.....	62
Table 4-8. Subscores for Mathematics by Grade.....	63
Table 4-9. Subscores for Science by Grade.....	64
Table 4-10. Summary Statistics for Classical Item Difficulties for ELA.....	67
Table 4-11. Summary Statistics for Classical Item Difficulties for Mathematics	67
Table 4-12. Summary Statistics for Classical Item Difficulties for Science	68
Table 4-13. Summary Statistics for Classical Item Discrimination for ELA.....	68
Table 4-14. Summary Statistics for Classical Item Discrimination for Mathematics.....	69
Table 4-15. Summary Statistics for Classical Item Discrimination for Science	69
Table 4-16. Summary Statistics for Item Response Theory Item Difficulty for ELA	70
Table 4-17. Summary Statistics for Item Response Theory Item Difficulty for Mathematics	70
Table 4-18. Summary Statistics for Item Response Theory Item Difficulty for Science.....	70
Table 4-19. Summary Statistics for Item Response Theory Item Discrimination for ELA	71
Table 4-20. Summary Statistics for Item Response Theory Item Discrimination for Mathematics	71
Table 4-21. Summary Statistics for Item Response Theory Item Discrimination for Science	71
Table 4-22. English Language Arts Cut Scores	74
Table 4-23. Mathematics Cut Scores.....	74
Table 4-24. Science Cut Scores	74
Table 4-25. English Language Arts, Mathematics, and Science Scaling Constants.....	75

Table 4-26. Number and Participation Rate (PR) of Enrolled and Tested Students by Subject and Grade	76
Table 4-27. Participation Rate by Demographic Characteristics and State Board of Education (SBOE) District	78
Table 4-28. Percentage of Tested Students by Demographic Characteristic and Grade	79
Table 4-29. Scale-Score Descriptive Statistics for English Language Arts	80
Table 4-30. Scale-Score Descriptive Statistics for Mathematics	80
Table 4-31. Scale-Score Descriptive Statistics for Science	80
Table 4-32. Percentage of Students Achieving at Each Performance Level (PL) for ELA, Mathematics, and Science	81
Table 4-33. English Language Arts Mean and Standard Deviation of Scale Scores by Grade and Student Subgroup.....	85
Table 4-34. Mathematics Mean and Standard Deviation of Scale Scores by Grade and Student Subgroup..	86
Table 4-35. Science Mean and Standard Deviation of Scale Scores by Grade and Student Group.....	87
Table 4-36. Total Number of Enrolled Students by Subject and Grade for 2021–2025	88
Table 4-37. Proficiency Rates for English Language Arts, 2021–2025	90
Table 4-38. Proficiency Rates for Mathematics, 2021–2025	90
Table 4-39. Proficiency Rates for Science, 2021–2025	90
Table 5-1. Accessibility Tools for KAP	94
Table 5-2. Accommodations Available for KAP Assessments.....	97
Table 5-3. Frequency of Accommodation Requests by Grade	98
Table 6-1. Panelist Count by Table Assignment and Grade	102
Table 6-2. Panel Composition	103
Table 6-3. Kansas Policy PLDs	104
Table 6-4. Agenda for Bi-grade Panels	105
Table 6-5. KAP ELA Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric)	109
Table 6-6. KAP Math Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric)	110
Table 6-7. KAP Science Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric).....	111
Table 6-8. Articulation Recommendations.....	113
Table 6-9. Results of the Standard-Setting Evaluation	114
Table 6-10. Results of the Articulation Panel Evaluation Across All Panels	115
Table 6-11. Kansas Assessment Scale and Cut Scores for the 2025 Revised State Assessments	118
Table A-1. Grade 3 English Language Arts (ELA) Summative Assessment Blueprint	123
Table A-2. Grade 4 English Language Arts (ELA) Summative Assessment Blueprint	123
Table A-3. Grade 5 English Language Arts (ELA) Summative Assessment Blueprint	124
Table A-4. Grade 6 English Language Arts (ELA) Summative Assessment Blueprint	124
Table A-5. Grade 7 English Language Arts (ELA) Summative Assessment Blueprint	125
Table A-6. Grade 8 English Language Arts (ELA) Summative Assessment Blueprint	125
Table A-7. Grade 10 English Language Arts (ELA) Summative Assessment Blueprint	126
Table A-8. Grade 3 Math Summative Assessment Blueprint	126
Table A-9. Grade 4 Math Summative Assessment Blueprint	126
Table A-10. Grade 5 Math Summative Assessment Blueprint.....	126
Table A-11. Grade 6 Math Summative Assessment Blueprint.....	127

Table A-12. Grade 7 Math Summative Assessment Blueprint.....	127
Table A-13. Grade 8 Math Summative Assessment Blueprint.....	127
Table A-14. Grade 10 Math Summative Assessment Blueprint.....	127
Table A-15. Grade 5 Science Summative Assessment Blueprint	128
Table A-16. Grade 8 Science Summative Assessment Blueprint	128
Table A-17. Grade 11 Science Summative Assessment Blueprint	128
Table B-1. Word Count and Lexile Range by Grade.....	129
Table C-1. Demographic Information of Grade 3–5 Content Passage Review Panelists.....	130
Table C-2. Demographic Information of Grade 6–8 Content Passage Review Panelists.....	131
Table C-3. Demographic Information of Grade 10 Content Passage Review Panelists	132
Table C-4. Demographic Information of Grade 3–5 Fairness Passage Review Panelists	133
Table C-5. Demographic Information of Grade 6–8 Fairness Passage Review Panelists	134
Table C-6. Demographic Information of Grade 10 Fairness Passage Review Panelists.....	135
Table D-1. Demographic Information of English Language Arts Content Item Review Panelists	136
Table D-2. Demographic Information of English Language Arts Fairness Item Review Panelists	138
Table D-3. Demographic Information of Mathematics Content Item Review Panelists	140
Table D-4. Demographic Information of Mathematics Fairness Item Review Panelists.....	141
Table D-5. Demographic Information of Science Content Item Review Panelists	142
Table D-6. Demographic Information of Science Fairness Item Review Panelists	143
Table E-1. Item Statistics Flagging Criteria.....	144
Table F-1. Demographic Information of English Language Arts Data Review Panelists	145
Table F-2. Demographic Information of Mathematics Data Review Panelists.....	146
Table F-3. Demographic Information of Science Data Review Panelists	147
Table G-1. ELA Subscore Reliability.....	148
Table G-2. Math Subscore Reliability.....	149
Table G-3. Science Subscore Reliability	150
Table H-1. Likert-Scale Questions (Percentages).....	151
Table H-2. Yes/No Questions (Counts Only)	152
Table I-1. Likert-Scale Questions (Percentages)	153
Table I-2. Yes/No Questions (Counts Only).....	153

Table of Figures

Figure 3-1. Grade 4 and Grade 8 ELA Proficiency Rate: 2025 KAP vs. 2024 NAEP	47
Figure 3-2. Grade 4 and Grade 8 Mathematics Proficiency Rate: 2025 KAP vs. 2024 NAEP.....	47
Figure 4-1. Test Information Function for English Language Arts	57
Figure 4-2. Test Information Function for Mathematics	58
Figure 4-3. Test Information Function for Science	59
Figure 4-4. Performance-Level Distribution for English Language Arts.....	82
Figure 4-5. Performance-Level Distribution for Mathematics.....	83
Figure 4-6. Performance-Level Distribution for Science.....	84
Figure 4-7. Participation Rates for 2021–2025 by Subject and Grade	89
Figure 6-1. Panelist Rating Example	107

Chapter 1 | Statewide System of Standards and Assessments

The Kansas Assessment Program (KAP), a program of the Kansas State Board of Education (SBOE), is mandated by the Kansas State Legislature. In addition, the English language arts (ELA), mathematics, and science components of KAP are used to comply with federal elementary and secondary education legislation. The three main purposes of KAP, as stated in the [2024–2025 KAP Test Coordinator Manual](#), are to:

- measure specific claims related to the Kansas Standards (KS) in grades 3–8 and high school.
- report individual student scores along with each student’s performance level.
- provide subscale and total scores that can be used with local assessment scores to assist in improving a building’s or district’s programs in ELA, mathematics, and science.

The state statutory authority behind KAP is Kan. Stat. Ann. §72-5170 (2024). According to this statute, the SBOE is mandated, in part, to

- design and adopt a school performance accreditation system based upon improvement in performance that reflects high academic standards and is measurable,
- establish curriculum standards that reflect high academic standards for the core academic areas of mathematics, science, reading, writing, and social studies, and
- provide for statewide assessments in the core academic areas of mathematics, science, reading, writing, and social studies and determine performance levels on the statewide assessments.

KAP provides summative assessment in ELA, mathematics, and science for all students in grades 3–8 and high school, except students with significant cognitive disabilities, who are eligible for alternate assessments. The 2024–2025 KAP assessments were developed from 2021 to 2024 and were first administered as the operational forms for 2025 summative assessments. The current technical manual provides description and analysis results for the 2024–2025 KAP assessment, including an introduction on content standards, a description of test development and test forms used for the 2025 assessment, technical analysis results using 2025 assessment data, a description of the process for setting the achievement standards, and a summary of validity evidence to support the interpretation of test scores for intended test uses.

1.1. State Adoption of Academic Content Standards for All Students

The 2017 Kansas Mathematics Standards were adopted by the SBOE in August 2017. More information about the 2017 Kansas Mathematics Standards can be found on the [KSDE Mathematics Standards website](#). For ELA, the SBOE adopted the current Kansas standards

in 2023. These standards are available [here](#). The 2025 KAP ELA assessments reflect the 2023 Kansas Standards for ELA, and the 2025 mathematics assessments reflected the 2017 Kansas Mathematics Standards.

For science, the SBOE adopted the Kansas Science Standards in 2013. In 2018, the Kansas science standards review committee reviewed the 2013 Kansas Science Standards and concluded that no updates to the standards were needed. More information about the 2013 Kansas Science Standards can be found on the [KSDE Science Standards website](#). The current 2025 KAP science assessments reflect the 2013 Kansas Science Standards.

1.2. Challenging Academic Content Standards

Committees of Kansas educators and stakeholders provided input on the standards in Kansas. These standards supported the vision of the KSDE: to lead the world in the success of each student (refer to the [Kansas State Board of Education web page](#)). These standards help schools equip students with the academic, cognitive, metacognitive, technical, and employability skills required for postsecondary success, as well as the capacity to positively affect the world around them. The Kansas Standards are Kansas’s coherent and rigorous academic content standards, which adhere to the SBOE’s mission. The mission of the Kansas State Board of Education is to prepare Kansas students for lifelong success through rigorous, quality academic instruction; career training; and character development according to each student’s gifts and talents.

1.2.1. Goals of Kansas Standards

As described in the [2023 Kansas English Language Arts standards](#): “The 2023 English Language Arts Standards were updated to reflect the vision of the Kansas State Board of Education: To Lead the World in the Success of Each Student. These standards have been reviewed to support each school district’s ability to utilize structured literacy as the explicit and evidence-based approach to teaching literacy skills to all students.”

There are five foundational practices included in the English Language Arts Foundational Practices, they are

1. Write, speak, read and listen appropriately in all disciplines.
2. Seek out and work to understand diverse perspectives.
3. Use knowledge gained from literacy experiences to solve problems.
4. Create multimodal versions of texts for a range of purposes and audiences.
5. Self-regulate and monitor growth in writing, speaking, reading and listening.

The 2017 Kansas Mathematics Standards were created to define what students should understand and be able to do in their study of mathematics. Mathematical understanding

is the ability to justify, in a way appropriate to the student’s mathematical maturity, why a particular mathematical statement is true or where a mathematical rule comes from. The student who can explain the rule understands mathematics and may have a better chance to succeed at a less familiar task. Mathematical understanding and procedural skills are equally important, and both are assessable using mathematical tasks of sufficient richness.

The 2013 Kansas Science Standards closely align with the Next Generation Science Standards (NGSS). The NGSS are based on the *Framework for K–12 Science Education* developed in 2012 by the National Research Council of the National Academies. However, the intent of the NGSS is to put the *Framework* into practice by coupling the practice with content, providing performance expectations while leaving curricular and instructional decisions to states and educators, and evaluating students on the degree of understanding of a full Disciplinary Core Idea.

1.2.2. Process and Timeline

The ELA standards were updated in 2023 to better align with the science of reading and structured literacy. After development, there was a window for public comments and review of the updated standards. The updated standards were presented to the SBOE in May 2023. The SBOE adopted the standards in June 2023.

The previous Kansas Mathematics Standards were reviewed, written, and edited by the Kansas mathematics standards writing and review committee between March 2016 and May 2017. Minutes of these meetings were kept, explaining the decisions that were made (KSDE, 2017). The committee presented the updated standards to the SBOE in July 2017 for adoption, followed by a window for public comments and review of the updated standards. In August 2017, the SBOE approved the adoption (KSDE, 2017).

The 2013 Kansas Science Standards were reviewed in 2018 by the Kansas science standards review committee. After reviewing, the committee concluded that no updates to the standards were needed. For the 2013 Kansas Science Standards, Kansas, as one the lead states in developing NGSS, had educators review the NGSS and make the recommendation of adopting these standards to the SBOE in 2013.

1.2.3. Standards Review Committees

To ensure that educators from across the state had an opportunity to nominate either themselves or someone else to serve on the standards review committees, information about the formation of the committees was distributed to the education community through email distribution lists, meetings, and the SBOE. Nominations were collected via a registration site that recorded the nominee’s demographic information, committee group of interest, years of work experience, and highest level of education. KSDE staff ensured that the standards review committees for ELA and science and the standards writing and review committee for mathematics consisted of diverse genders, races, ethnicities, and teaching levels (K–12 and postsecondary), and that every state district was represented.

Each committee also included an ad hoc group that consisted of representatives from various educational organizations, business communities, and KSDE, as well as legislators, parents, and other community members. Although the ad hoc group members participated in discussions during the standards review process, they did not provide an official vote on the final product that was subsequently reviewed and adopted by the SBOE (KSDE, 2017). As for the 2013 Kansas NGSS review committee, 60 members from across the state participated, comprised of K–12 science educators, postsecondary science professors, and business and industry professionals.

1.3. Required Assessments and Intended Population

KAP measures student achievement in the subject areas of ELA, mathematics, and science. The subject areas and grades tested are as follows:

- ELA in grades 3–8 and 10,
- mathematics in grades 3–8 and 10, and
- science in grades 5, 8, and 11.¹

1.4. Policies for Including All Students in Assessments

Kansas is committed to including all students in the KAP assessment. Students enrolled in Kansas public schools must take either the KAP assessment or the alternate assessment. Additionally, students designated as English learners must also take the English language proficiency test. In the first year entering the United States, English learners are required to take the KAP mathematics and science tests. They are not required to take the ELA assessment but must take the Kansas English Language Proficiency Assessment (KELPA). In their second year in Kansas schools, English learners are required to take KAP or alternate assessments in ELA, math and science when appropriate.

Eligible students with significant cognitive disabilities, typically no more than 1% of Kansas students, take the Dynamic Learning Maps® (DLM®) alternate assessment for ELA, mathematics, and science. Other students with Individualized Education Programs (IEPs), 504 plans, or Student Intervention Team plans take the KAP assessment but can use accommodations consistent with their Personal Needs and Preferences (PNP) Profile. A PNP Profile is a piece of information in a student’s educational file that describes the accommodations provided to students during instruction. If unapproved accommodations are used (e.g., reading aloud to a student on the KAP ELA test), the student test record is considered invalid. A detailed summary of accommodations for KAP can be found in Chapter 5 | Inclusion of All Students.

¹ A grade 11 science test was administered during 2025, but scores were not reported due to misalignment between student ability and test difficulty.

Exemptions from KAP assessments are granted to students who, during the testing window,

- move into a different school that is not in the same school district,
- experience catastrophic illnesses or accidents,
- are serving long-term suspension,
- are truant for more than two consecutive weeks and have filed truancy paperwork,
- are incarcerated in an adult facility, and/or
- are in a special detention center.

Chapter 2 | Assessment System Operations

The development of any test requires many critical decisions regarding, for example, the content and cognitive complexity, the appropriate scope of that content for subject areas, and the number of items associated with each test. These decisions are not made in isolation but must be reasonable across all grade levels of the assessment. Together, these decisions guide the test construction process and products.

2.1. Test Design and Development

The Achievement and Assessment Institute (AAI) worked with the Kansas State Department of Education (KSDE) to determine the content to be assessed by Kansas Assessment Program (KAP) tests for each subject area and grade over multiple years. The 2025 English language arts (ELA) summative assessment aligned with the 2023 Kansas Standards for ELA, while the math summative assessment aligned with the 2017 Kansas Standards for Mathematics. The 2025 science summative assessment aligned to the 2013 Kansas Science Standards. All 2025 KAP assessments feature updated blueprints, and every item was newly developed to align with the current Kansas Standards. Table 2-1 outlines the test development timeline for KAP assessments.

Table 2-1. Development Timeline for KAP Assessments

Milestone	Date
Blueprint development	2020–2022
Passage and item development	2021–2024
Field testing items	Spring 2022–2024
Test construction	2024–2025
First operational administration aligned to new blueprint	Spring 2025
Standard setting	Summer 2025

Note. Newly developed items were field tested on previous summative assessments.

2.1.1. Test Blueprints

The blueprints were developed in collaboration among the AAI content team, KSDE, and educators from winter 2020 to fall 2022. In November 2020, KSDE indicated that the priority of Kansas competencies should inform the content priority of the blueprints. KSDE and AAI began the blueprint design process by assigning the priority levels (high, medium, and low) of standards based on the priority of competencies. This assignment also considered standard vertical articulation across the grades to ensure there was reasonable progression of those standards. Then, item point values were assigned to the standards based on their designated priority levels, forming the foundation of the test blueprints. These blueprints encompass all assessed areas within the standards. In general, higher-priority standards receive more score points, reflecting their greater instructional and assessment emphasis. Rather than specifying the number of items, the blueprints define score-point ranges to allow for deeper assessment of content within individual items and to provide flexibility in test form construction.

To determine appropriate test length, KSDE considered the balance between total testing time and the number of subscore categories. With input from the Technical Advisory Committee (TAC), KSDE established a minimum of 40 score points per summative assessment across all subjects and grade levels. Kansas educators were then consulted to review the blueprints for reasonableness and equitable access to content. Based on their feedback, the blueprints were revised accordingly.

Appendix A | Test Blueprints shows the test blueprints for ELA, mathematics, and science. The test blueprints for ELA and science include reporting categories that are fairly consistent across grade levels. For ELA, the main components of the blueprint are writing, reading: literary, and reading: informational. For science, the main components of the blueprints are Physical Science, Life Science, and Earth & Space. Proportions of each reporting category under these main components vary by grade. Mathematics, on the other hand, includes varying reporting categories across grades that reflect the topical nature of math standards. Each KAP assessment item is worth 1 or 2 points. Generally, 2-point items are scored as 0, 1, and 2 points and require more complex cognitive processes.

In ELA, all items are passage-based. The blueprint specifies text complexity requirements for the passages, categorized into four levels: slight, slight-moderate, moderate-complex, and complex. At each grade level, the assessment includes one to two passages with slight complexity, three to four passages with slight-moderate and moderate-complex texts, and one to two passages with complex texts. Section 2.2.1. English Language Arts Passage Selection and Review outlines the process to determine text complexity level.

The mathematics assessment consists of both block and standalone items. Block items are grouped around a shared stimulus and address multiple standards, while standalone items—whether mathematical or real-world in context—are each aligned to a single standard and do not share a stimulus.

Similarly, the science assessment includes both block-based and standalone items. Each item is anchored in a real-world science phenomenon within the domain of Physical Science, Life Science, or Earth & Space Science. Both item types may assess a single standard or integrate multiple standards.

2.1.2. Test Design

In 2025, all three subject areas employed a fixed-form test design. All assessment items underwent and passed comprehensive item reviews, and the resulting test forms were designed as operational. Each subject had one operational form administered across two sessions. To deter cheating, each session included multiple blocks of items that were identical in content but presented in different orders. Research indicates that item order does not affect item performance (Hohensinn et al., 2011; Li et al., 2012), so blocks with reordered items were considered equivalent forms.

Students without Personal Needs and Preferences (PNP) settings were randomly assigned

to one of the standard test forms, while a designated test form was provided for students requiring accommodations. As described in Section 2.1.1. Test Blueprints, all operational forms have items totaling a minimum of 40 score points. Table 2-2 shows the number of operational items on each KAP assessment session by subject.

Table 2-2. Number of Operational Items on the 2025 KAP Assessment by Subject and Session

Subject	Grade	Items			Score Points		
		Session		Total	Session		Total
		1	2		1	2	
ELA	3	22	18	40	22	19	41
	4	24	16	40	26	16	42
	5	22	18	40	23	18	41
	6	20	20	40	20	22	42
	7	21	19	40	24	19	43
	8	22	18	40	22	19	41
	10	20	20	40	21	22	43
Math	3	20	20	40	20	20	40
	4	20	20	40	20	20	40
	5	20	20	40	20	20	40
	6	20	20	40	20	20	40
	7	20	20	40	22	20	42
	8	20	20	40	20	20	40
Science	5	20	20	40	21	26	47
	8	20	20	40	22	20	42

Note. ELA = English language arts.

2.1.3. Operational Test Construction

The 2025 test forms of all grades and subjects were constructed using the following procedures and guidelines:

- Items and passages were approved by KSDE after passing all item reviews for inclusion in assessments.
- Test forms were assembled following the content specifications in the blueprint; items with the best psychometric characteristics were preferentially selected.
 - Items with negative or very low discrimination, or extremely low or high item difficulty, were not selected.
 - Each test form included a wide range of item difficulties, and the average difficulty was of a moderate level.
- Test forms were reviewed to eliminate item enemies (e.g., items that might clue answers to other items).

- Test forms were reviewed and approved by psychometric staff for psychometric properties.
 - This was to ensure that test forms include items with a wide range of item difficulties and with a moderate level of test difficulty.
 - This also served to ensure that test sessions have enough items with high item discrimination.
- For math grade 6 and above, calculators are introduced in the assessments. Within the Session 1 form, calculator-inactive items are presented to students first, followed by calculator-active items. Aside from grade 6, Session 2 forms include all calculator-inactive items. The grade 6 Session 2 form included two sections, first calculator-inactive items and then calculator active items

After content staff and psychometricians identified items for form construction, an accessibility expert reviewed the inclusion and functionality of accessibility features to ensure the widest range of students can access the items. The accessibility enhancement incorporated knowledge of disabilities (e.g., blind or low vision, deaf or hard of hearing) and language learners, as well as checking for overall task clarity and cohesion. Accessibility features that were incorporated into items include:

- accessible color palettes
- appropriate color contrast settings
- alternative text on images
- keyboard navigation
- compatibility with commonly used assistive technology products, such as screen readers
- braille
- key word translations (English to Spanish) for math and science forms
- American Sign Language (ASL) videos
- text-to-speech

2.2. Item Development

Item development entailed various efforts to ensure item quality, including ongoing research into best practices and new item types, developing and using subject-area item specifications, updating materials for item writer training, recruiting new or additional item writers, conducting item writer training for new item writers or refresher training for continuing item writers, conducting various internal and external item reviews, and updating item review processes and trainings.

In addition to internal processes, before appearing on any assessment, items were reviewed by educator content reviewers, educator fairness reviewers, and KSDE staff. The AAI content team used item review feedback to revise test items as needed. Items were then prepared for field testing according to test specifications and established guidelines for both general and accommodated presentations such as text-to-speech. After field

testing, educators reviewed items from an item statistics perspective during data review. Items that did not pass any review were either discarded or revised and put through the item review process again. Items that did pass all reviews were eligible for inclusion on the operational assessment.

The next sections describe item development for the 2025 KAP assessments. ELA, math, and science item writing occurred from 2020 to 2023. Items were written and reviewed at different times. However, item writing and review procedures were consistent across all subject areas and grades.

2.2.1. English Language Arts Passage Selection and Review

For ELA, the passage selection process started with identifying appropriate passages from public domain works, published books, and websites, as well as commissioning original passages from test development vendors and internal staff. Passage development took place between summer 2020 and summer 2023. Both passages and items underwent a rigorous development process.

Passage Text Complexity

Assessment passages included commissioned, permissioned, and public domain readings. AAI's content team used a combination of quantitative and qualitative resources to guide grade placement and evaluate text complexity.

Quantitative Analysis

Quantitative measures—including word count, Flesch Kincaid, Coh-Metrix, and Lexile value—were used to determine the grade placement. Each grade level had established ranges for word count and Lexile value (Appendix B | Word Count and Lexile Range). Lexile ratings are familiar to most educators because many instructional publishers provide Lexile ranges for their materials. These measures reflect sentence length and vocabulary difficulty. Lexile values were obtained using the Lexile Analyzer, an online tool.

The grade placement process followed these steps:

- Determine whether the passage falls within the Lexile range for the target grade level.
- If the Lexile value falls outside the grade-level range, assign the passage to the appropriate grade based on the Lexile result.
- Once placed, modify the passages as needed to meet the grade-level word count range.

Qualitative Analysis

Qualitative measures were then applied to determine the text complexity levels. Two separate rubrics—one for informational texts and one for literary texts—were used by AAI content specialists and KSDE staff.

The informational-text rubric evaluates passages across four text complexity levels using the following criteria:

- purpose
- text structure (organization, text features, use of graphics)
- language features (conventionality, vocabulary, sentence structure)
- knowledge demands (subject-matter knowledge and intertextuality)

The literary-text rubric also rates passages on four complexity levels, using these criteria:

- meaning
- text structure (narration, order of events, use of graphics)
- language features (conventionality, vocabulary, sentence structure)
- knowledge demands (life experiences, cultural knowledge, intertextuality, subject-matter knowledge)

The appropriate rubric was applied based on passage type. Reviewers first read the passage, including any introductory text, and then evaluated each criterion within the rubric. All components were considered holistically to determine the overall text complexity level.

KSDE determined final grade placement for each passage based on rubric ratings and overall text complexity.

Internal Passage Review

Passages from all sources underwent multiple rounds of internal review, including content, accessibility, graphics (if applicable), and editorial reviews. AAI content experts and accessibility specialists first reviewed passages for content accuracy and accessibility issues (e.g., inaccurate or outdated information). Passages were then fact-checked by a professional editor.

The passage review rubric covered four areas: (1) craft and structural considerations; (2) words, sentences, and paragraphs; (3) representation, accessibility, and factual accuracy; and (4) language complexity. Each area included guiding questions to help content and accessibility reviewers evaluate passages both holistically and analytically.

Graphics and editorial reviews ensured that any graphics met design specifications, adhered to the style guide, and complied with accessibility standards.

Passages accepted during internal review were presented to KSDE prior to external educator reviews. External passage reviews were conducted five different times: summer 2021, winter 2022, summer 2022, winter 2023, and summer 2023. The summer 2021, 2022, and 2023 reviews were combined with item reviews. Table 2-3 provides details on the exact timing, location, and number of panelists for each external review.

External Passage Review

All external panels were grade-banded (grades 3–5, 6–8, and high school). For some reviews (summer 2022 and summer 2023), not all grade bands reviewed passages. Each grade band consisted of two panels: a content review panel and a fairness review panel. Both panels reviewed all passages within their respective grade bands. All external passage reviews followed the same procedure and used the same review criteria.

Table 2-3. External Passage Review Time, Location, and Panelist Counts

External Passage Review	Time	Location	Number of Content Review Panelists			Number of Fairness Review Panelists		
			G3–5	G6–8	G10	G3–5	G6–8	G10
Summer 2021	June 4–29	Virtual	7	6	7	6	8	7
Winter 2022	Jan. 24 – Feb. 11	Virtual	7	8	7	7	5	6
Summer 2022	July 25–29	Virtual	5	—	—	6	—	—
Winter 2023	Feb. 1–3	Salina, KS	4	6	6	6	8	7
Summer 2023	July 17–21	Salina, KS	6	6	—	7	6	—

Note. “—” indicates no review conducted.

AAI and KSDE staff recruited Kansas educators to serve as passage reviewers for two separate types of reviews: content review and fairness review. Prospective external passage reviewers completed an online survey in which they indicated their demographic information, teaching experience, professional qualifications, content expertise, experience with the standards, and experience with special education or English learners (ELs). Based on survey results, AAI compiled a list of potential reviewers and provided it to KSDE for final panel selection. Demographic information for the content and fairness passage review panels by grade band are summarized in Appendix C | Demographic Information for Passage Reviewers.

The content review panels consisted of classroom teachers representing different SBOE districts. Most content review panelists had more than 10 years of experience and backgrounds in English instruction within the grade band they reviewed. The fairness review panelists were also composed mainly of classroom teachers from multiple SBOE districts. About half of the educators had fewer than 10 years of experience, while the other half had more. Fairness panelists brought expertise in supporting diverse student populations, including students from various demographic groups, students receiving special education services, ELs, and students experiencing trauma.

Each external passage review included three phases: orientation, independent review, and facilitated discussion. For virtual reviews, the orientation and discussions were conducted via an online meeting platform. For in-person item reviews, orientation and independent review occurred at a secure review facility. In both formats, reviewers completed independent reviews in a secure online system. After orientation, reviewers evaluated passages asynchronously at their own pace and submitted feedback by the designated

deadline. The facilitated panel discussion then provided an opportunity to review and consolidate feedback from individual reviewers.

During the orientation, panelists received security protocols and background information about the standards, the purposes of the passages and items, and the test development cycle. They were also briefed on the distinct purposes of the two review types:

- Content passage review ensures passages are grade-level appropriate, accurate, engaging, and aligned with the intended standards.
- Fairness passage review ensures passages are accessible and appropriate for all Kansas students.

Panelists were trained to use KAP’s secure online review system to record feedback and were provided contact information for support during the asynchronous portion.

Following orientation, reviewers independently evaluated passages and entered feedback into the system. Content reviewers considered grade-level appropriateness, text complexity, and the standards potentially supported by each passage. They ensured that

- ideas progress logically,
- paragraphs are well structured with topic sentences and supporting details,
- vocabulary is grade-level appropriate,
- sentence length and variety are appropriate,
- content matches the intended complexity level,
- language demands are suitable for the grade,
- topics are engaging, developmentally appropriate, meaningful, and current.

Fairness reviewers examined passages for potential bias and sensitivity issues, ensuring that each passage

- does not present barriers preventing any student group from understanding the text,
- avoids portraying any group unfavorably or stereotypically,
- excludes offensive or emotionally charged language or symbolism, and
- maintains accessibility without diluting rigor.

After independent reviews, panels reconvened to discuss feedback collectively. Notes for potential revisions—such as adjusting character names, clarifying definitions, or reducing redundancy—were documented during the discussion.

Following the external reviews, the AAI content team compiled all feedback and discussion notes and presented them to KSDE. KSDE staff then rated each passage as “accepted,” “requires revision,” or “rejected.” Passages accepted or revised to KSDE’s satisfaction advanced to item development. Finalized passages were approved by KSDE before item writing began.

2.2.2. Item Writing

All math and science items and half of the ELA items on the KAP summative assessment were written by True North Education Consultants via a contract. The True North Education Consultants wrote items in multiple stages.

- Stage 1 occurred in 2020–2021 and included items for ELA (grades 3–8 and 10), math (grades 3–8 and 10), and science (grades 5, 8, and 11).
- Stage 2 occurred in 2021–2022 and included items for ELA (grades 3–8 and 10), math (grades 3–8 and 10) and science (grades 5 and 8).
- Stage 3 occurred in 2022–2023 and included science grade 11.
- The rest of the ELA items were written internally at AAI in the spring of 2023 and the summer of 2023.

All item writers from True North held at least a bachelor’s degree, and most had 15 or more years of teaching experience. Internal AAI item writers were either full-time staff or contract employees, all with classroom teaching experience, prior item development experience, and familiarity with educational standards. All item writers received training on the Kansas ELA standards, alignments, and best practices in item writing.

Both vendor and internal item writing followed the same process. To ensure vendor writers were familiar with KAP item specifications, AAI provided resources such as style and graphics guidelines, item templates, and content standards. The item development process included four main steps: item writing, content review, editorial review, and a final quality check against item writing guidelines.

During item writing, writers created items tagged with relevant metadata (e.g., grade, content area, Depth-of-Knowledge [DOK] level, correct key or scoring guide, and distractor rationales). Writers also specified graphics requirements and calculator usage. For items requiring graphics, writers submitted requests to the graphics team and reviewed the completed graphics for accuracy, adherence to specifications, and compliance with accessibility standards.

2.2.3. Item Review

The item review process involved several stages:

1. Internal content and accessibility review
2. Internal editorial review
3. KSDE review
4. External review (content and fairness) using multiple panelists
5. Data review using multiple panelists

Internal Item Review

During the internal content review, reviewers evaluated each item, scoring method, Depth-of-Knowledge (DOK) level, alignment to standards, and associated graphics. Reviewers

examined the item’s content, structure, format, clarity, key accuracy, and graphical quality. Key guiding questions included:

- What does the item ask? Is it important and not trivial?
- Does it clearly align to the standard?
- Is the key the only possible key?
- If it is a technology-enhanced item with more than one correct key, are the correct answer options the only defensible correct answer options?
- Is the item complete (e.g., with content codes, key or keys, grade, scoring, DOK, and test program identified)?
- Are distractors viable, and do they represent common errors and misconceptions?
- Are the item content and approach appropriate for the designated grade?
- Is the readability appropriate for the designated grade?
- As a set, do the items cover the blueprint?

After the AAI content team performed the internal content review, the items were then reviewed by the accessibility and editorial teams, in an iterative process. Internal accessibility reviews of assessment content focus on ensuring all test takers can demonstrate their knowledge without unnecessary barriers. The review process examines whether content is perceivable across different presentation formats (screen readers, braille, etc.), uses clear and grade-appropriate language free from jargon or cultural assumptions, represents diverse groups equitably without reinforcing stereotypes, and avoids emotionally charged topics that could distract from the actual assessment objectives. The goal is to ensure assessments measure the intended knowledge and skills without being impacted by construct-irrelevant factors like accessibility barriers or emotional reactions.

Following the completion of the internal reviews, the items were presented to KSDE. KSDE reviewed the items through a holistic lens to ensure the items were ready to present to external reviewers.

For editorial review, the editor consulted and ensured compliance with the house style guide. The editor also considered item integrity, item format, item content and language, possible ambiguity, key security, and issues related to bias, sensitivity, and accessibility. This editorial level of review involved, but was not limited to, the following actions:

- Eliminate confusion or vague wording.
- Ensure consistency and coverage of various responses according to scoring guides.
- Suggest rewrites as necessary (final wording approved during resolution of items or item set).
- Ensure consistency of usage and terms.
- Check for correct grammar and usage; fix typographical and spelling errors.
- Review for common structural, logical, and design issues (e.g., outlier answer choices, nonparallel choice phrasing, key security).

Then, the item writer and editor resolved each element of the review. For the final item review, the content reviewer examined the final item and checked the following topics:

- General issues
 - The question clearly addresses the standard.
 - All content is accurate (graphics, passage, and question).
 - No economic, cultural, ethnic, gender, or religious bias is present.
 - Context is realistic.
 - Context and reading level are grade appropriate.
- Universal Design for Learning (UDL)
 - Wording is clear and concise.
 - Syntax uses present tense and active voice when possible.
 - Reading level is as low as possible.
 - Simple sentence structure (subject-verb-object) is used when possible.
 - Sentences are concise.
 - No colloquialisms or words with double meanings are used unless that is the target of assessment.
- Item stems
 - Stem presents a definite, explicit, and singular question.
 - Stem is brief and free of irrelevant information.
 - Stem includes appropriate qualifiers (e.g., best, most likely) if necessary.
 - Stem is worded positively when possible.
- Item options
 - All choices are plausible.
 - Distractors capture common misconceptions or errors.
 - Numerical options are in ascending or descending order.
 - Answer choices are grammatically parallel (e.g., same part of speech, same sentence structure).
 - All choices are grammatically consistent with the stem.
 - There are no clues to the correct answer (e.g., opposites, antonyms, synonyms, phrases repeated from the stem).
 - All choices contain the same level of detail.
 - Answer choices are the same length or exhibit parallelism in their lengths.

External Item Review

This section describes the reviewers and review process for two external reviews: content review and fairness review. Items passing the two external reviews were then field tested, and student-response data were used in item analyses. For items that were tagged for atypical psychometric functioning, the item analysis results were presented to external data review panelists. This section also describes the reviewers and review process for data review.

For ELA, external item reviews occurred five different times: summer 2021, summer 2022, winter 2023, summer 2023, and fall 2023. Table 2-4 includes the exact time, location, and

number of panelists of different panels for each ELA external item review. All panels were grade-banded: grades 3–5, grades 6–8, and grade 10.

For math, external item reviews occurred in two different times: summer 2021 and summer 2022. Table 2-5 includes the exact time, location, and number of panelists of different panels for each math external item review. All panels were grade-banded: grades 3–5, grades 6–8, and grade 10.

For science, external item reviews occurred in three different times: summer 2021, summer 2022, and summer 2023. Table 2-6 includes the exact time, location, and number of panelists of different panels for each science external item review. Each grade had its own panel: grade 5, grade 8, and high school. For summer 2021 and summer 2022 science item review, items reviewed were grade 5 and 8 items. For summer 2023 science item review, items reviewed were grade 11 items.

For the three subjects, each grade or grade band consisted of two panels: a content review panel and a fairness review panel. Both the content and fairness panels reviewed all items at that grade or grade band. All external item reviews followed the same procedure.

Table 2-4. ELA External Item Review Time, Location, and Panelist Counts

External Item Review	Time	Location	Number of Content Review Panelists			Number of Fairness Review Panelists		
			G3–5	G6–8	G10	G3–5	G6–8	G10
Summer 2021	June 4–29	Virtual	7	6	7	6	8	7
Summer 2022	July 25–29	Virtual	5	4	—	6	5	—
Winter 2023	Feb. 1–3	Salina, KS	—	—	6	—	—	7
Summer 2023	July 17–21	Salina, KS	6	6	8	7	6	6
Fall 2023	Nov. 13–17	Virtual	6	6	—	6	5	—

Note. “—” indicates no review conducted.

Table 2-5. Math External Item Review Time, Location, and Panelist Counts

External Item Review	Time	Location	Number of Content Review Panelists			Number of Fairness Review Panelists		
			G3–5	G6–8	G10	G3–5	G6–8	G10
Summer 2021	June 4–29	Virtual	6	6	7	6	6	7
Summer 2022	July 25–29	Virtual	3	3	3	5	3	3

Table 2-6. Science External Item Review Time, Location, and Panelist Counts

External Item Review	Time	Location	Number of Content Review Panelists			Number of Fairness Review Panelists		
			G5	G8	G11	G5	G8	G11
Summer 2021	June 4–29	Virtual	6	4	—	3	6	—
Summer 2022	July 25–29	Virtual	4	3	—	4	4	—
Summer 2023	July 17–21	Lawrence, KS	—	—	6	—	—	6

Note. “—” indicates no review conducted.

External Item Reviewers

AAI and KSDE staff recruited Kansas educators to serve as item reviewers for two separate types of reviews: content review and fairness review. Prospective external item reviewers for ELA, math, and science completed an online survey in which they indicated their demographic information, teaching experience, professional qualifications, content expertise, experience with the standards, and special education or EL status. After reviewing the survey results, AAI staff provided the list of potential reviewers to KSDE staff for selection for each panel. The demographic information for the content item review panels and for the fairness item review panels for different grades / grade bands and subjects are summarized in Appendix D | Demographic Information for External Item Reviewers.

The content review panels consisted of classroom teachers from different SBOE districts. Most content review panelists had more than 10 years of experience. Moreover, content panel educators included educators with backgrounds in teaching English. Most fairness review panelists were classroom teachers from different SBOE districts. About half of the educators had fewer than 10 years of experience, and the other half had more than 10 years of experience. Fairness panels included educators with experience teaching students with various backgrounds, including students in different demographic subgroups, special education students, ELs, students experiencing trauma, etc.

Content review panels for ELA and math were organized by grade band, while science panels were organized by grade. In general, the majority of panelists were female, White, non-Hispanic classroom teachers or district staff, comprising more than half of all participants. Some reviewers served on multiple panels, as their content expertise often extended across grade levels. Bias and sensitivity panels were also convened, including members from diverse backgrounds to reflect the demographic and cultural diversity of Kansas. All item reviews were conducted through a secure online system. After completing web-based training, reviewers evaluated items independently and submitted feedback by the assigned deadline.

External Item Review Process

The external reviews included an orientation, an independent review of the items, and a panel discussion of the items after the independent review. For the virtual item review, the orientation and panel discussions occurred via an online meeting platform. For the in-

person item review, the orientation and independent reviews occurred in a secure review facility. For both formats, the independent review occurred through a secure online reviewing system. After completing the orientation, reviewers evaluated items at their own pace and provided feedback. Then, the panels discussed all feedback and possible revisions for items during the synchronous discussion.

All item reviewers participated in an orientation for either the content review or fairness review. Orientations included two components: one session of specialized training for content or fairness review and another session for the online review system. Both the content and fairness orientations included security reminders, background information, and an overview of the assessment development process. After the orientation, panelists were encouraged to ask questions about their review responsibilities. Finally, panelists engaged in a practice set of items to practice item review and familiarize themselves with the review platform. This opportunity allowed for review and discussion of several items to ensure panelists were comfortable and confident in the expectations for the review. After panelists confirmed their confidence with the review and system, they engaged in the independent review and provided feedback.

Content reviewers considered every aspect of each item: alignment to content standards, appropriateness (i.e., content, context, and vocabulary for the grade and subject), correct and incorrect answers, and utility and clarity of graphics or stimulus.

In general, content reviewers checked items for

- alignment to standards,
- grade appropriateness, including language, context, direction, and item type,
- content errors,
- alignment to Depth-of-Knowledge level,
- key errors, and
- alignment to the Standards for Mathematical Practice (math only).

For the fairness review, panelists independently reviewed items and submitted their feedback. Fairness reviewers identified barriers not related to content that could prevent students from demonstrating what they know and can do. These barriers included unfamiliar language, linguistic complexity, potentially sensitive topics, stereotypes (both positive and negative), accessibility for special populations, and issues with cultural or experiential knowledge.

In general, fairness reviewers checked items to

- identify potential bias and sensitivity issues,
- ensure all items were appropriate and accessible for all Kansas students, including
 - principles of UDL incorporated into items, and
 - appropriate language complexity for all students, and
- ensure representation that broadly and generally reflected the student population.

During the panel discussion, the panelists discussed their feedback. Panelists were encouraged to provide any additional feedback they had on the items during discussion. The facilitators captured the discussion notes via the online reviewing system.

Following the content and fairness external item reviews, the AAI content team combined panelists’ feedback and discussion notes with items and presented this information to KSDE to review. KSDE provided ratings (accept, additional revisions required, or reject) on the items. Then the AAI content team modified the items that were deemed by KSDE as requiring minimal edits and revisions. Items that had been edited or revised were presented to KSDE for final sign-off. Once the items had secured final sign-off, they were ready for field testing.

Data Review

Following the field-test administration of the items, statistical analyses of the items were conducted using classical item statistics. Items with statistics that fell outside criteria were flagged and reviewed in data review. Item statistical flagging criteria are explained in Appendix E | Item Statistics Flagging Criteria.

For ELA, data reviews occurred three different times: summer 2022, summer 2023, and summer 2024. Table 2-7 includes the exact time, location, and number of panelists of different panels for each ELA data review. All panels were grade-banded: grades 3–5, grades 6–8, and grade 10. For some ELA item reviews (e.g., summer 2023), not all grade bands were included.

For math, data reviews occurred two different times: summer 2022 and summer 2023. Table 2-8 includes the exact time, location, and number of panelists of different panels for each math data review. All panels were grade-banded: grades 3–5, grades 6–8, and grade 10.

For science, data reviews occurred three different times: summer 2022, summer 2023, and summer 2024. Table 2-9 includes the exact time, location, and number of panelists of different panels for each science data review. Each grade had its own panels: grade 5, grade 8, and grade 11. For the summer 2022 and summer 2023 science data reviews, items reviewed were grade 5 and 8 items. For the summer 2024 science data review, items reviewed were grade 11 items. All data reviews followed the same procedure.

Table 2-7. ELA Data Review Time, Location, and Panelist Counts

Data Review	Time	Location	Number of Data Review Panelists		
			G3-5	G6-8	G10
Summer 2022	July 25–29	Virtual	11	9	—
Summer 2023	July 17–21	Salina, KS	6	6	—
Summer 2024	June 10–21	Virtual	5	6	4

Note. “—” indicates no review conducted.

Table 2-8. Math Data Review Time, Location, and Panelist Counts

Data Review	Time	Location	Number of Data Review Panelists		
			G3-5	G6-8	G10
Summer 2022	July 25-29	Virtual	6	6	8
Summer 2023	July 17-21	Lawrence, KS	6	6	6
Summer 2024	June 10-21	Virtual	6	4	4

Table 2-9. Science Data Review Time, Location, and Panelist Counts

Data Review	Time	Location	Number of Data Review Panelists		
			G5	G8	G11
Summer 2022	July 25-29	Virtual	8	7	—
Summer 2023	July 17-21	Lawrence, KS	5	5	—
Summer 2024	June 10-21	Virtual	5	4	4

Note. “—” indicates no review conducted.

Data Reviewers

AAI and KSDE staff recruited Kansas educators to serve as data reviewers. Prospective data reviewers ELA, math, and science completed an online survey in which they indicated their demographic information, teaching experience, professional qualifications, content expertise, experience with the standards, and experience with special education or EL students. After reviewing the survey results, AAI staff provided the list of potential reviewers to KSDE staff for selection to each panel. The demographic information for the data review panels for different grades / grade bands and subjects are summarized in Appendix F | Demographic Information for Data Reviewers. Note that in the summer 2022 data review, the educators who participated in the data review also did the external item review.

Data Review Process

The data reviews included an orientation and a panel review of the items. For the virtual data review, orientation and panel reviews occurred via an online meeting platform. For the in-person item review, orientation and panel reviews occurred in a secure review facility. After completing the orientation, reviewers evaluated items based on item statistics and provided feedback together.

During the orientation, the panelists were presented with security reminders, the purpose of data review, background information regarding the items, the process to review the items, introduction on item statistics and flagging criteria, and additional information regarding the location of materials and instructions on how to supply their feedback. Also, the facilitator had a practice set of items to calibrate the expectations of the review. Additionally, a one-page description of the statistics, their meaning, and guiding questions were provided to the panelists for guidance during their review.

After the orientation, the facilitator led the data review by presenting each flagged item and asked for the initial feedback to see if there was anything about the item that stood out from a content or fairness perspective. Then, the facilitator explained the statistical flags associated with the item and gathered feedback on the rationale from panelists that could account for the flags. The co-facilitator took notes during discussion.

For the in-person data review, the panelists were provided with an orientation that included the same information as the virtual panels. Then the reviewers reviewed the items independently to identify any content or fairness issues with the items. After the independent review, the facilitator led the reviewers through the flagged items, explained the data flags, and asked reviewers to identify any content rationales for why the items may have been flagged. During this discussion, reviewers could ask questions about item performance, statistical data, and typical response patterns, as well as any other questions that arose regarding the item performance.

Following the virtual and in-person data reviews, the AAI content team combined panelists' feedback, discussion notes, and item statistics with the items and presented this information to KSDE to review. KSDE provided ratings (accept, revise, or reject for operational use) on the items. The accepted items were ready for constructing the summative assessment test forms. The AAI content team revised any items that were deemed by KSDE as revisable and planned to re-field test these revised items. KSDE had the final disposition on all items flagged during data review, as well as final sign-off on items that were revised during the reconciliation process.

2.3. Test Administration

Large-scale assessment requires a standardized test administration process to prevent the unintended effects of administration differences. The standardized test administration procedures are described in the [KAP Summative Test Administrator Manual \(TAM\)](#). The TAM provides information regarding standardized test administration for districts, schools, and teachers. It also provides guidance on the administration procedure for the 2024–2025 KAP assessment. Main topics of the TAM include

- overview of the KAP summative assessment,
- test security and ethics,
- administration guidelines,
- accommodations,
- preparation activities before test administration,
- directions for test administration on testing day,
- activities for after test administration, and
- resources for test administration.

For all subjects, grades, and students, KAP is entirely computer based, and the delivery platform is Kite® Student Portal. The Kite Suite is a platform used to create, administer, monitor, score, and report on various types of online assessments. The electronic item

bank, online administration system, and student responses are stored in the Kite platform. To take KAP assessments, Student Portal must be installed on students' testing devices. The 2025 KAP summative testing window opened on Monday, March 24, 2025, and closed on Friday, April 25, 2025. The KAP summative remote testing window opened on April 28, 2025, and closed on Friday May 9, 2025. Each test session was designed to take approximately one class period (i.e., 45–60 minutes). Thus, one test was designed to take approximately two class periods. However, all KAP tests are untimed, as enough time should be given to students to finish testing.

2.3.1. Test Administration and Security Training

Kansas uses a train-the-trainer model, in which district test coordinators (DTCs) receive training directly from KSDE and then train building-level personnel before local test administration. First, the test administration and security training for all Kansas DTCs include: test coordinator regional test security and ethics training held in different locations in September and January, and test coordinator virtual training webinars. Then, DTCs train local test administrators.

The regional trainings were in-person trainings that covered test coordinator responsibilities, test security and ethics, using Kite, and updates for the upcoming year. After training, DTCs had to verify their training and agree to adhere to policies and practices covered in the training. For the 2024–2025 test administration, all DTCs needed to attend and verify the training before January 15, 2025. The test security and ethics training covers test security procedures, test administration monitoring, roles and responsibilities, reporting testing discrepancies and potential violations, reporting item issues, the security of personally identifiable information (PII), accommodations, and appropriate and inappropriate testing practices. DTC virtual training webinars were held on the third Tuesday of each month in 2024–2025. DTCs who could not attend live webinars could access online training materials and webinar recordings at any time on the KAP website through the [DTC Virtual Training Webinars](#) link. The trainings were offered by AAI in partnership with KSDE and provided updates on KAP and Kite technology, an overview of important training dates, a description of accommodations, directions for ordering braille booklets, and updates on special circumstance codes.

For local training, DTCs train staff members who administer state assessments at the district or building level before testing begins. Local staff members include administrators, educators, paraeducators, and other appointed staff members. The training includes test security and ethics, reporting, and accommodations. For all training at the district and building level, DTCs document the personnel, time, and method of the training, and maintain records at the district and building levels. Anyone administering a KAP assessment had to complete all district- and building-level training by March 21, 2025. After completing training, staff administering state assessments signed an agreement to abide by state ethical testing practices and provide written verification.

2.3.2. Test Administration Procedures

The TAM includes guidelines for administering KAP assessments in a standardized and secure procedure; KSDE developed and approved these guidelines. All test administrators are required to read the TAM. The standardized and secure test administration procedures before, during, and after KAP administration are described in the next sections, and more detailed information can be found in the TAM. Further details about administration-related accommodations can be found in Chapter 5 | Inclusion of All Students in this manual.

Before KAP Administration

Before KAP administration, local testing windows should be scheduled to ensure all students can finish testing before the end of the school day and before the end of the testing window. Once the local testing windows are scheduled, those dates should be added in Kite Educator Portal, a platform to manage user and student data. Districts then register students for testing using the Kansas Individual Data on Students (KIDS) system. Student records are inserted into Kite Educator Portal and accounts created for use in Kite Student Portal. Also, teachers should complete the PNP Profile settings for students who need accommodations and enter special circumstance codes for students who cannot take KAP assessments.

As the local testing window nears, test administrators should

- prepare the room for testing (e.g., remove instructional materials that may give clues),
- have appropriate manipulatives for the math and science assessments,
- be familiar with rules for using resource sheets and calculators for math assessments,
- have students' individual usernames and passwords ready,
- have access to daily access codes (DACs) needed to access KAP assessments, and
- have needed materials ready (e.g., pencils, scratch paper, clocks, and headphones).

To better prepare students for KAP assessments, educators should strongly encourage students to take the practice tests. The [Kite Practice Test Guide](#) is available for educators to support students' access to practice tests. Kite Student Portal provides practice tests to help students gain confidence navigating assessments and become familiar with different KAP item types before taking the test.

There are two types of practice tests: a technology practice test that includes various item types using simple content, and a subject-oriented practice test that features various item types with subject-oriented content. The subject-oriented practice test also provides a deeper look at different tools. All practice tests are grade-banded: the technology practice test includes grade bands K–1, 2–5, and 6–12; and the subject-oriented practice test includes grade bands 3–5, 6–8, and 10–11. Technology practice tests are not secure and should be used to help students gain experience taking assessments on the online

platform (i.e., Kite Student Portal) and feel confident taking the actual KAP assessments.

During KAP Administration

On assessment day, test administrators make sure students are taking the correct test, help students log in to Kite Student Portal, instruct students to enter the DAC, and remind students not to disrupt others if they finish early. After testing starts, test administrators should only read specific scripts provided in the TAM as instructions. To ensure a quiet testing environment and help students focus on testing, test administrators also need to follow proctoring guidelines given in the TAM during testing. Last, before a student exits the test, test administrators should verify the review screen to ensure all items were answered.

After KAP Administration

After one KAP test session administration, test administrators should collect all materials, such as manipulatives for math and science. Also, test administrators should collect and destroy all materials, including test tickets and scratch paper. Braille forms are sent back to Assessment & Technology Solutions (ATS) for secure storage. Then, DTCs monitor student testing status and reactivate student testing sessions if needed.

2.4. Monitoring Test Administration

Test administration monitoring includes monitoring both testing and testing data. Testing monitoring also includes both local monitoring and KSDE visits. For local monitoring, district and building-level users (DTCs and BTCs) can use the Kite Educator Portal to track test session completion, including progress at the individual item level. DTCs and other test administrators are responsible for identifying any testing discrepancies; testing discrepancies are any violations of standard test administration procedures. After testing discrepancies are identified, superintendents or their designees are responsible for reporting them in writing.

Every year during the testing window, KSDE staff and members of the Kansas Assessment Advisory Council (KAAC) visit approximately 5–10% of Kansas schools to monitor administration and test security. The monitoring visits were held during the testing window between March 24, 2025 and April 25, 2025. A total of 12 districts, including 23 buildings and 26 classes, were monitored in person. These districts include two districts that did not complete test security training on time and five districts that volunteered to be monitored. Fifteen observers, including KSDE staff and members of the KAAC, participated in in-person monitoring. All observers used the monitoring checklist in the [KSDE Test Security Guidelines](#) to evaluate testing sessions.

Virtual testing uses the live proctoring service provided by MonitorEDU to monitor the testing activities during administration. For each test session, a professionally trained proctor from MonitorEDU conducts a live proctoring using an online meeting tool. Before testing starts, the proctor does a 360° room scan through the camera and checks whether any smart devices, such as smart glasses or watches, are being used. Once testing starts,

the proctor monitors the behavior of the student to detect any irregularities and provides immediate intervention if necessary.

The testing and data monitoring for non-virtual windows in 2025 was the same as in previous years. During testing-data monitoring, no irregularities were found for the non-virtual window. However, for the virtual window, 18 students were flagged by MonitorEDU. Two of the 18 students were identified as having invalid test results.

2.5. Test Security

AAI implemented a series of policies and procedures to ensure the security of test results and prevent test irregularities. The majority of test-related activities happen on online platforms. The electronic item bank, online administration system, and student responses are stored in the Kite platform, which is designed and maintained by the AAI technology team. Three portals were designed within the Kite platform to serve different needs:

- Content Portal, for item and test development
- Educator Portal, for educators to input and access test and student information
- Student Portal, for online testing

The Kite platform uses Amazon Web Services (AWS) in high-availability mode with no single point of failure. Using AWS ensures that the loss of any given server or even of an entire availability zone (i.e., data center) will have minimal impact on Kite platform availability. Recovery times are very short, ranging from no downtime (for loss of most servers) to a few minutes (for loss of an entire data center). Moreover, AWS fully manages the recovery, which runs in high-availability mode and is automatic. Using a service provided by AWS, the Kite platform has a multilayered design to prevent denial-of-service attacks and system intrusion. The Kite platform moved to AWS in 2017. Since then, the Kite platform has experienced no outages that have affected testing.

KAP used [Caveon](#), a test security company, to detect any exposure of test content on social media or elsewhere on the internet. Caveon Web Patrol leverages both automated technologies and human analysis to find and track threats to the assessment program, such as illicit discussion, distribution, and sale of test content online. It monitors various platforms, including “brain-dump” sites,² test-prep sites, online auctions, and social media. Regular reports categorize threats by risk level and provide actionable recommendations. Caveon maintains a log of potential risks and uses a secure platform, Caveon Core, to notify clients of high-risk threats. They collaborate with state agencies to remove infringing content, starting with cease-and-desist letters and escalating as needed.

² “Brain-dump” sites are websites or platforms that provide unauthorized or unethical access to exam questions and answers, often in the form of exact replicas or memorized content from real exams. These sites typically gather and share exam content that is meant to be confidential, allowing users to “dump” what they remember from an exam for others to use. Using brain-dump sites is considered cheating.

Timely notifications and frequent client collaboration are key to their success.

In 2025, before the test window started, all test content was shared with Caveon. Then, Caveon monitored various websites and social media sites during both virtual and non-virtual testing windows until one week after the virtual window closed. Caveon reported exposed items were taken down from social media platforms within 48 hours. There were a total of 219 exposures encompassing 99 unique items. Item 3008174 had 12 exposures, the most of any item. Grade 10 had the most exposure, at 162 exposures encompassing 60 unique items. Math was the most impacted subject, with 210 exposures. The most used platform was Gauthmath, with 202 uploads. There were 204 items uploaded to sites as photos, and 16 were typed into sites.

ATS concluded that exposure was minimal and recommended that none of the items be replaced. Given the limitations of the sites used, exposures were likely limited to the individual students uploading them. Since those students will not take the assessment again, the exposure risk is mitigated. ATS will monitor these items in future administrations and see if the item difficulty decreases over time—possibly indicating items are searchable prior to being removed from the platforms.

2.5.1. Prevention of Test Irregularities

Standardized test administration procedure is the key way to prevent test irregularities. KAP offers test administration and security training to ensure standardized test administration procedures. The training introduces appropriate and inappropriate testing practices to help test administrators understand test irregularities. Also, all test administrators must sign an agreement to abide by state ethical testing practices. For detailed information about test administration and security training, please refer to Section 2.3.1. Test Administration and Security Training.

Moreover, KSDE has predetermined procedures to deal with testing discrepancies and possible security violations. All testing discrepancies and possible security violations should be reported to KSDE. Upon breach of security, appropriate consequences are put in place at the district level. Depending on the uniqueness of each case, possible steps vary and may include, but are not limited to,

- no action because the breach was not severe enough to warrant any action,
- KSDE action, such as a written letter or phone call to the superintendent or DTC, stating concerns and monitoring action steps,
- retesting of students,
- removal of test proctors from testing rooms, and/or
- follow-up visits by a KSDE monitor the next testing year to ensure changes to inappropriate practices have been made.

For more details, refer to the [KSDE Test Security Guidelines](#).

2.5.2. Detection of Test Irregularities

The procedures for detecting test irregularities can be separated into in-person monitoring and data monitoring. For the details of in-person monitoring, please refer to Section 2.4. Monitoring Test Administration.

During the operational window, monitoring of testing data was conducted by the AAI technology team and the AAI psychometric team. The technology team conducted data validation daily to monitor system usage and identify testing irregularities. System usage includes a DTC training log, click history of student responses, test-taking hours, test-status summary, server load, the number of Kite Service Desk (i.e., support for Educator Portal and Student Portal) tickets, and the frequency of test reactivations. Testing irregularities include fast test-taking behavior (i.e., students finished a test section in a short amount of time), irregular testing time (i.e., a test session started or ended outside of school hours), tests reactivated by users (i.e., test administrators) or by the system, and student enrollment or demographic data errors.

The AAI psychometric team periodically conducted student-response data checks to ensure quality administration. Those checks included verifying that

- student demographic information was entered in Kite,
- student test information values were accurate,
- students received only one score for each item,
- item scores matched the possible item scores,
- all possible item scores were obtained by at least 1% of students,
- each student had only one set of demographic information,
- each student took only one test form in each subject,
- distributions of demographic information and test information were reasonable,
- students' raw scores in the first session had a strong relationship with those in the second session, and
- the frequency distribution of students' raw scores was smooth, was bell-shaped, and generally increased and then decreased as raw scores increased.

2.5.3. Remediation of Test Irregularities

DTCs and other test administrators are responsible for identifying any test irregularities. Superintendents or their designees are responsible for reporting any identified test irregularities to KSDE in writing. After consulting with staff from KSDE, districts need to implement appropriate consequences. For the details of possible consequences, refer to the [KSDE Test Security Guidelines](#).

For test irregularities at the individual student level or class level, affected students will need to be retested. For test irregularities at the state level, such as the breach of assessment items, the breached items will be excluded from scoring.

2.6. Protecting Data Integrity and Privacy

Protecting data integrity and privacy focuses on several important facets: test materials, test-related data, PII, and accommodation-related security. Data integrity and privacy should be protected through the whole testing cycle, from test development and administration to scoring and reporting. Moreover, to protect the data integrity and privacy of all facets, both physical security and online, platform security requirements should be met, and strict procedures should be in place during administration and reporting. The electronic item bank, online administration system, and student responses are stored in the Kite platform. Refer to Section 2.5. Test Security for details about the process used to ensure Kite security.

2.6.1. Test Materials Integrity

To protect the integrity of test materials during test development, the physical security requirements are met by using hosting providers that conform to the Statement on Auditing Standards (SAS-70) for physical access and Payment Card Industry Data Security Standard (PCI DSS) compliance. Most activities related to project management, test development, and data analysis take place at AAI. Both centers are in secure wings that can be accessed only with a key or key card. In general, work is done either at one of our centers using secure server systems or a secure virtual private network connection. Moreover, the electronic item bank stored in Kite Content Portal can be accessed only by authorized users of Kite. For any activities involving external item reviewers, such as Kansas educators, all participants are required to sign nondisclosure agreements to ensure item and task confidentiality and security.

To protect the integrity of test materials during test administration, specialized training and certification for test administrators are required (described in Section 2.3.1. Test Administration and Security Training). Test administrators are expected to deliver assessments with integrity and to maintain the security of assessments. State, district, and school users are expected to complete the security agreement in Educator Portal each year. By accepting the security agreement, users agree to not store or save assessment materials to computers or personal storage devices, to not print assessment materials, to not share personal passwords with others, and to destroy assessment materials after administration.

2.6.2. Test-Related Data Integrity and Privacy

For test administration, all Kite portals handle educator and administrative passwords using industry-standard encryption techniques. Users must create strong passwords and may change their own passwords at any time in accordance with the password policy. All portals generate access records that system administrators can review to track access. Access to individual Kite portals is controlled according to established policies for that application and the data it maintains. All access policies and accounts are reviewed periodically to ensure that access to systems is limited to the appropriate populations.

DTCs attend the test security and ethics training provided by KSDE and oversee test security for the entire district. They establish procedures that determine which appropriate personnel can access Educator Portal and their role assignments within the district. DTCs also remove or deactivate from Educator Portal any users who leave the district or change roles within a district. Moreover, DTCs establish and describe processes ensuring the usernames and passwords in Educator Portal are exclusive to the users, and also confirm that users' rights are permitted according to their roles.

2.6.3. Personally Identifiable Information Integrity and Privacy

In accordance with the Family Educational Rights and Privacy Act (FERPA) and the principle of least privilege, students, teachers, operators, and administrators are granted access only to the specific student records for which they have a legitimate educational requirement. All users receive the minimum level of access necessary to perform their duties. Throughout each school year, security levels, groups, and access are reviewed periodically to ensure continued compliance.

All test administrators are informed that PII should not be conveyed when reporting testing issues. The documentation for Kansas regarding allowable identifiers in an email specifies that only the State Student Identifier and no other identifying details (e.g., name, district, school) should be provided in an email. In cases when the Kite Service Desk needs to be contacted, students' PII cannot be sent via email or live chat.

For scoring and reporting purposes, students' PII data are stored on secure servers in AAI. AAI staff who work on KAP materials and who have access to PII data on servers are required to complete annual KSDE information technology security and data privacy training to ensure compliance with FERPA. Operational access to all secure servers is controlled by keys that are provided only to system administrators in the operations team that manages the production data center. Access to networking equipment and hardware consoles is limited to the data center itself; remote access to these devices is limited to the data center administration host.

After scoring and reporting are complete, ATS provides student assessment data (e.g., return files, score reports) to KSDE. Those data are placed on a secure drive that only specific members of the ATS and KSDE teams can access. For school and district reporting, scores from more than 10 students are needed for aggregated results; this is to prevent identifying individual students' scores. Descriptions of KAP results in technical documentation are reported only at the aggregated level.

2.6.4. Accommodations-Related Data Integrity and Privacy

Local staff members who administer a state assessment must complete the test administration and security training given by DTCs, sign an agreement to abide by state ethical testing practices, and provide written verification of training before local testing begins. The training covers the ethics of testing, test security, and reporting and documenting accommodations. To ensure security related to accommodations, DTCs

need to establish procedures for entering student accommodations in the PNP Profile in Educator Portal and keep documentation for text-to-speech accommodations and other accommodations that require deviating from general administration of the assessment. More information about selecting and entering information in the PNP Profile is in Section 5.4.1. Selection of Accommodations. Text-to-speech accommodation of ELA passages must be approved by KSDE before testing. Thus, either DTCs or BTCs need to submit the need for text-to-speech accommodation of ELA passages to KSDE at the beginning of the year. During the assessment, Kite audio (i.e., via headsets) is used for text-to-speech accommodation, rather than a human reader.

Chapter 3 | Technical Quality: Validity

According to the *Standards for Educational and Psychological Testing* (the *Standards* hereafter), *validity* refers to “the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests” (American Psychological Association [APA] et al., 2014, p. 11).

The *Standards* (APA et al., 2014) provide a framework for describing the sources of evidence that should be considered when evaluating test score validity. These sources include evidence based on test content, response processes, internal test structure, relationships between test scores and other variables, and the consequences of testing. The validation process involves the ongoing collection of a variety of evidence to support the proposed test score interpretations and uses. This chapter mainly describes aspects of the Kansas Assessment Program (KAP) assessments that support KAP test score interpretations and uses.

Because the purpose of validity evidence is to support the intended uses of test scores, it is necessary to identify the purposes of a test before providing evidence to support test validity. The purposes of the KAP assessment, as described at the beginning of this manual, include (a) measuring specific claims related to the Kansas Standards, (b) reporting students’ academic performances, and (c) using the KAP assessment with local assessment scores to assist in improving educational programs in the three subject areas.

The gathered evidence on test content, response process, and internal structure supports the use of the KAP assessment to measure the Kansas Standards as defined in the test blueprints. Information on test reliability, fairness and accessibility, and scoring and scaling justify the use of KAP test scores for reporting students’ academic performances. Validity evidence from other sources, such as comparing KAP results with National Assessment of Educational Progress (NAEP) results, uses additional data to validate the use of KAP test scores.

3.1. Validity Evidence Based on Test Content

Validity evidence based on test content refers to how well test content related to specific content domains matches with what the test was intended to measure. Content evidence for KAP assessments comes from the alignment between KAP items and the Kansas Standards, from the congruence between the test and the test blueprint, and from the congruence between the test blueprint and the standards (i.e., a balance of representation of standards). Achievement and Assessment Institute (AAI) content specialists used the following procedural steps to evaluate the content validity of the KAP assessment.

- Collaborating with KSDE, develop the test blueprint and specification, and evaluate the relationship between the blueprint and the Kansas Standards.
- Conduct internal content and fairness reviews of KAP items.
- Conduct content reviews of KAP items using a panel of content experts to see

whether the items measure the intended construct or whether sources of construct-irrelevant variance exist.

- Conduct fairness reviews of KAP items to avoid bias and sensitivity issues related to specific subpopulations.
- Refer to KSDE content experts to validate the content.
- Evaluate the alignment between KAP assessments and the Kansas Standards.
- Evaluate the degree to which the assessment addresses the depth and breadth expectations of the Kansas Standards in terms of the blueprint.

Chapter 2 | Assessment System Operations presents validity evidence related to the development of the test blueprint, item and test development, and item review. As described in those chapters, the KAP blueprint has the same structure as the Kansas Standards. Test content specialists developed and aligned all KAP items with the Kansas Standards, and item development followed well-established procedures. After item development, items underwent multiple rounds of content and bias reviews. After field-test administration, test content specialists reviewed the items' statistical properties, evaluating items from content and psychometric perspectives before their selection for operational use. Districts then administered KAP assessments according to standardized procedures and provided accommodations for students with special needs. The following list summarizes the effort to ensure content validity.

- The development of the blueprint is a collaborative process among AAI, the Kansas State Department of Education (KSDE), and educators in Kansas. The blueprint uses the same framework as the Kansas Standards, which ensures the range and variety of standards measured in KAP are appropriate as indicated in Section 2.1.1. Test Blueprints.
- The proportion of items for each reporting category or claim in Appendix A | Test Blueprints shows that each reporting category or claim has adequate score points to represent the knowledge and skills described in the Kansas Standards.
- AAI and KSDE selected and trained qualified item writers to ensure they would write high-quality items.
- AAI and KSDE established detailed item development guidelines to train item writers. The item writers also participated in guided item writing.
- AAI content specialists and editors reviewed each new item and considered the grade-level appropriateness, graphics, grammar and punctuation, language demand, and distractor reasonableness.
- External content reviewers reviewed each item to make sure all items aligned with the Kansas Standards; they also considered grade-level appropriateness, checked the correct answer, evaluated the incorrect answers, evaluated the need for any included graphic or stimulus, and commented on the utility and clarity of graphics and stimuli.
- External bias, fairness, and sensitivity reviewers reviewed items to identify barriers that may prevent students from demonstrating what they know and can do when those barriers are not related to the content standards.

- Before items were selected for operational use, both AAI psychometricians and content leads reviewed the items’ classical item analysis and distractor analysis results to prevent items with extreme statistics from being used on operational forms.
- AAI accessibility experts reviewed each new item to make sure the widest range of students can access the items.
- Standardized administration of KAP assessments minimizes the effect of the variations in administration and provides accommodations for students who need them. Students are given ample time to complete the tests to avoid speediness issues.

3.2. Validity Evidence Based on Internal Structure

As described in the *Standards* (APA et al., 2014), *internal-structure evidence* refers to “the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based” (p. 13). Three sets of validity evidence about the internal structure provide (a) evidence that the KAP assessment is essentially unidimensional, (b) evidence that the item response theory (IRT) model used for each subject showed good fit results, and (c) evidence that the test contains no or few items flagged for significant and large differential item functioning (DIF), which helps support comparable measurement across groups.

3.2.1. Dimensionality

We applied confirmatory factor analysis (CFA) in order to evaluate whether a model with one dominant dimension fit the data reasonably well when the IRT scale was set. We carried out CFA using tetrachoric or polychoric correlations for binary or ordinal item responses and robust weighted least-squares estimation with the lavaan R package (Rosseel, 2012). The one-factor CFA model was considered to fit well if the comparative fit index (CFI) and Tucker Lewis Index (TLI) were 0.95 or greater and the root mean square error of approximation (RMSEA) was 0.05 or smaller.

Overall, all ELA, mathematics, and science tests may be reasonably treated as unidimensional. For ELA grades 3–8 and grade 10, the CFI and TLI were consistently 0.99 or higher, and RMSEA values ranged from 0.01 to 0.02, indicating excellent fit. For mathematics grades 3–8 and grade 10, the CFI and TLI values ranged from 0.98 to 0.99, with RMSEA values between 0.02 and 0.04. These results meet the fit criteria, supporting unidimensionality. For science grades 5, 8, and 11, the CFI and TLI values were between 0.95 and 0.99, and RMSEA values ranged from 0.01 to 0.04. Although grade 5 science was at the lower bound of the 0.95 threshold for CFI and TLI, the model still met the predefined criteria.

3.2.2. Item Response Theory and Model Assumptions

We analyzed KAP items using IRT. IRT is an industry standard for item analysis in large-scale K–12 assessment programs because of its item- and person-invariance claims.

However, IRT has several model assumptions that need to be fulfilled: model fit, local independence, and item-parameter invariance. The resulting inferences from any application of IRT depend on the degree to which the underlying assumptions are met.

This section describes the IRT models and calibration procedures used for all subjects and grades. The evaluation of IRT assumptions is presented as evidence of score validity.

Item Response Theory Calibration

We used IRT to calibrate item parameters to create a scale for each subject and grade. The 2025 test administration is the first operational administration with new test blueprints; therefore, a new KAP score scale is established. The following subsections introduce the IRT models, the sample used for calibration, the psychometric software, and the calibration procedures used for KAP.

Item Response Theory Model

We applied the two-parameter logistic (2PL) model (Birnbbaum, 1968) and the graded response model (GRM) (Samejima, 1969) to dichotomous and polytomous scored items, respectively. The choice of these two models contributed to the consistent and coherent interpretation of item parameters, as the 2PL is a special case of GRM that handles dichotomous items. The 2PL model defines the probability that a student of proficiency θ will answer item i correctly (u) as

$$P(u_i = 1|\theta) = \frac{e^{[a_i(\theta - b_i)]}}{1 + e^{[a_i(\theta - b_i)]}}, \quad (\text{Equation 3-1})$$

where a_i is the discrimination parameter, and b_i is the difficulty parameter.

Discrimination indicates how well the item distinguishes between students with higher or lower levels of proficiency; *difficulty* is the degree of item difficulty and is on the same scale as theta.

Under the GRM, the probability that u_i is equal to any observed score category v equals the cumulative probability of scores 0 to $v - 1$, minus the cumulative probability of scores v to the maximum score. The probability that the score is v or higher is

$$P(u_i = v|\theta) = \frac{e^{[a_i(\theta - b_{iv})]}}{1 + e^{[a_i(\theta - b_{iv})]}}, \quad (\text{Equation 3-2})$$

where a_i is the discrimination parameter, and b_{iv} is the difficulty parameter for score category v .

One discrimination parameter is estimated for each item; this parameter may be interpreted as the strength of association between the item and theta. For m response categories, there are $m - 1$ GRM b parameters. The b for category v is interpreted as the point on theta where the probability of scoring in category v or higher is 0.5.

Calibration Sample

Student responses available in the Kite system prior to April 27, the conclusion of the brick-and-mortar test administration window, were used for calibration. The calibration sample included all students who completed at least half of the items in each session and formally exited the test, excluding students who needed accommodations. Items that were omitted—that is, presented but left unanswered—were scored as incorrect (coded as 0). Table 3-1 provides the number of students by subject and grade for the sample used in IRT scale-setting calibration.

Table 3-1. Sample Size for Scale-Setting Calibration by Subject and Grade

Subject	Grade	Calibration Sample Size
ELA	3	33,563
	4	32,848
	5	33,842
	6	30,768
	7	31,433
	8	31,429
	10	31,376
Mathematics	3	33,955
	4	33,049
	5	33,961
	6	31,533
	7	31,824
	8	31,866
Science	10	33,182
	5	34,299
	8	32,098

Software

The mirt package (Chalmers, 2012) in R was used for IRT model estimation. The item-parameter calibration used the expectation-maximization algorithm. For all subjects and grades, the IRT calibrations converged; that is, the log-likelihood changes were smaller than 0.0001.

Calibration Procedures

For each subject and grade, a single-group concurrent calibration was conducted to place all item parameters onto the same scale for each subject and grade assessment. To accomplish this, we compiled all operational items of the same subject and grade into one file to create a student-by-item data matrix, which was then analyzed using estimation software for calibration.

IRT Model Evaluation

The validity inferences from the IRT results depend on the degree to which assumptions of the models are met and on how well the models fit the data. This section describes how the assumptions about IRT model fit, local independence, and item-parameter invariance are evaluated. All operational items were included in model evaluations for ELA, mathematics, and science.

Model Fit

We used the Q1 chi-squared (χ^2) fit statistic to evaluate the model fit for individual items during calibration, computed using the *mirt* package in R. The Q1 χ^2 fit statistic followed the χ^2 distribution with degrees of freedom defined by $(k - 1)(g - 1)$, where k is the number of item response categories and g is the number of θ -score groups used in forming the expected frequencies. Because χ^2 tests are highly sensitive to sample sizes, statistical significance alone may not indicate item misfit. To address this, we also evaluated item fit using Cramér's V (Cramér, 1946)—which summarizes the magnitude of deviation between observed and expected response frequencies. Practical interpretation followed common IRT residual-fit conventions: values below .05 were considered negligible, values between .05 and .10 suggested mild deviation, and values larger than .10 indicated notable misfit deserving review. Items were flagged when they showed significant Q1 results ($\alpha = .01$) and exhibited Cramér's V values of .10 or greater.

Items were flagged for potential model-fit issues if they showed statistically significant Q1 results at the $\alpha = 0.01$ level and exhibited medium to large effect sizes on Cramér's V .

Item removal (from calibration) was guided by a holistic approach, incorporating both statistical indicators and content requirements. Specifically, decisions regarding item removal were based on a combination of statistical flags and blueprint coverage, ensuring that both psychometric evidence and content requirements were considered. This process typically involves multiple rounds of item removal, recalibration, and reevaluation. In the final round, we identified five items that have model-fit issues: three items in math and two items in science. These five items have low discrimination (a parameter) and large standard errors of item parameter estimates. Because of the blueprint coverage issue, their discrimination parameter estimates were fixed to a more reasonable value of 0.25. These items will be monitored and item parameters updated in future administrations.

3.2.3. Differential Item Functioning

DIF analysis evaluates items for potential bias and examines whether an item shows statistical difference between two groups of students, controlling for student ability. We used logistic regression to detect items with uniform DIF (i.e., an item is consistently more difficult for one group of students than the other group for all ability levels). We used the Jodoin and Gierl (2001) DIF classification criteria to indicate the degree of DIF (i.e., negligible, moderate, large). When the DIF test is significant, large DIF is identified by a Nagelkerke R^2 change greater than or equal to 0.070; moderate DIF has a Nagelkerke R^2

change between 0.035 and 0.070; and negligible DIF has a Nagelkerke R^2 change is less than 0.035.

For each subject and grade, we examined DIF across gender (i.e., female vs. male), race (i.e., Black vs. White), and English learner (EL) status (i.e., EL vs. non-EL). For all subjects and grades, no items were flagged from among all operational items on three subjects for moderate or large gender-related DIF, or race-related DIF or EL status-related DIF. All results suggested that the item development process and procedures effectively addressed potential bias and sensitivity issues during the development phase.

3.3. Validity Evidence Based on Relations to Other Variables

As described in the *Standards*, “evidence based on relationships with other variables provides evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations” (APA et al., 2014, p. 16).

The evidence described above refers to external evidence and is categorized by two types: convergent and discriminant. When examining the relationship between students’ performance on different assessments that are intended to measure similar constructs, there is *convergent evidence*. When examining the relationships between students’ performance on different tests that are intended to measure different constructs, there is *discriminant evidence*. The correlations among different KAP subject scores are considered discriminant evidence. The comparisons of the KAP, NEAP, and Kansas English Language Proficiency Assessment (KELPA) performance are considered convergent evidence.

3.3.1. Relationships Among KAP Subjects

Discriminant evidence can be evaluated using the correlations between subjects, such as ELA and mathematics. Past studies showed high correlations between subjects, which indicates that subjects share some common traits; however, the correlations should not be too high. Table 3-2 shows the correlations and disattenuated correlations (correcting for measurement error) between subjects of the same grade, with values that range from 0.69 to 0.78 for correlations, and from 0.86 to 0.96 for disattenuated correlations. The lowest correlations among subjects are between ELA and mathematics, while the highest correlations are between ELA and science. After correcting for measurement error, the lowest disattenuated correlation is still between ELA and mathematics, and the highest disattenuated correlation is between ELA and science. According to Cohen (1988), a correlation larger than 0.50 represents a large effect size. Therefore, all observed correlations among KAP subjects reflect large effect sizes, indicating substantial—but not excessive—overlap among the constructs measured by each subject area.

Table 3-2. Correlations (C) and Disattenuated Correlations (DC) Among ELA, Mathematics, and Science Scores

Grade	ELA vs. Mathematics		ELA vs. Science		Math vs. Science	
	C	DC	C	DC	C	DC
3	0.73	0.86	—	—	—	—
4	0.72	0.86	—	—	—	—
5	0.73	0.86	0.76	0.96	0.73	0.90
6	0.74	0.90	—	—	—	—
7	0.71	0.90	—	—	—	—
8	0.74	0.89	0.78	0.96	0.76	0.88
10	0.69	0.90	—	—	—	—

3.3.2. Relationships Between KAP Assessment and National Assessment of Educational Progress

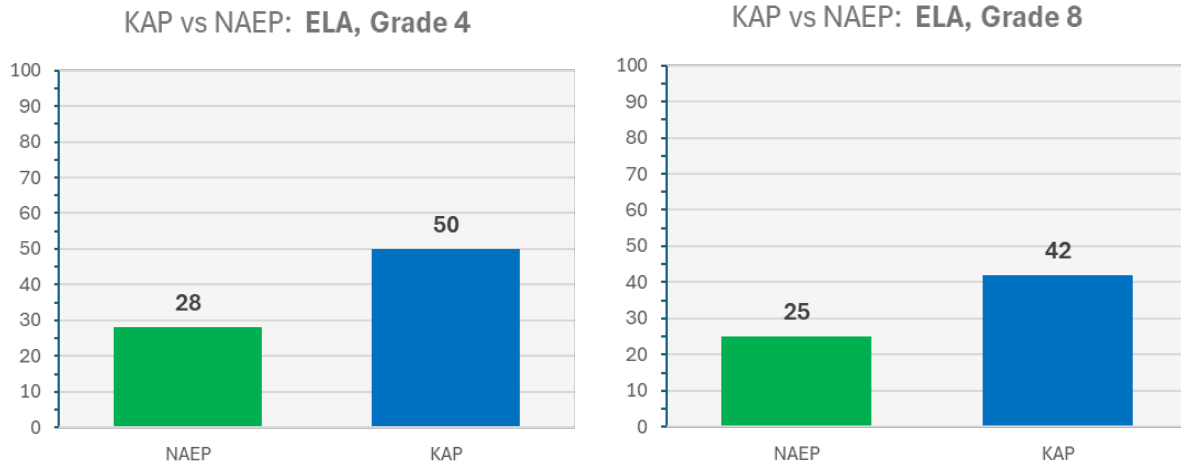
Convergence evidence requires an evaluation of the relationship between the target test with another test that measures a similar construct. The state of Kansas participates in the NAEP, also known as the Nation’s Report Card. NAEP is the largest nationally representative assessment of what American students know and can do, and it serves a different role than state assessments. NAEP assessments allow each state to be compared to national results and to evaluate progress over time. The results inform the public about the academic achievement of elementary (grade 4) and secondary (grade 8) students in Kansas and in the United States on ELA and mathematics.

Thus, the relationship between KAP and NAEP performance can be used as one source of convergent evidence. However, individual NAEP scores are not available. Only the trend of proficiency rates across years are compared between the two assessments. Although this is the first year for the refreshed KAP assessment and score scale, we plan to monitor the trends of proficiency rates across years. The trends of the two assessments can indicate the actual performance of Kansas students based on the two assessments measuring a similar construct. KSDE provides more information about NAEP on the [KSDE website](#).

KAP categorizes student performance by four performance levels (levels 1, 2, 3, and 4). The proficiency rate of KAP is the percentage of students in levels 3 and 4. NAEP categorizes student performance by three performance levels (Basic, Proficient, and Advanced). The proficiency rate of NAEP is the percentage of students in the Proficient and Advanced levels. Figure 3-1 and Figure 3-2 compare the 2025 KAP and 2024 NAEP proficiency rates for ELA and mathematics grades 4 and 8.³ We plan to monitor the biannual results of NAEP proficiency and compare them with the annual KAP assessment results.

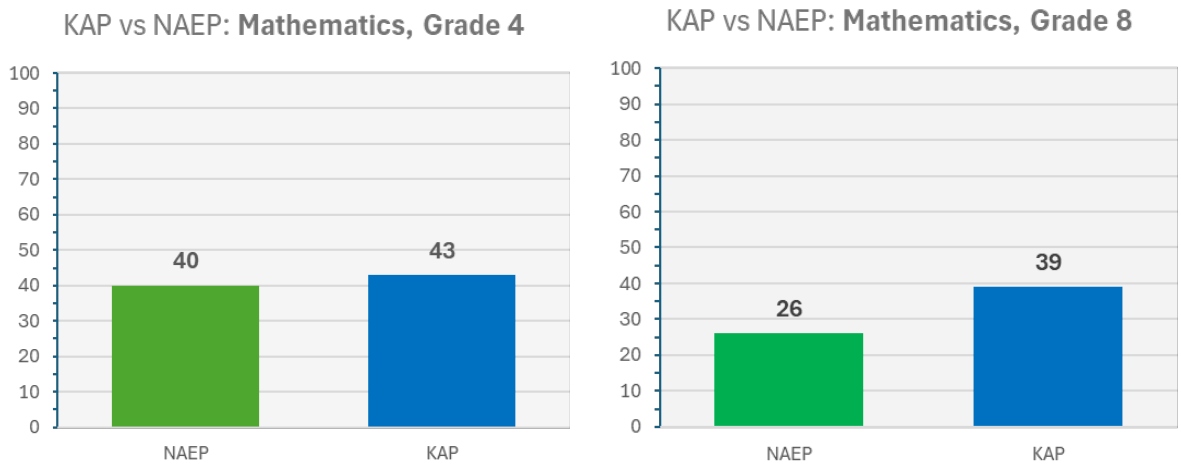
³ NAEP is administrated in odd-numbered years only. The 2021 NAEP assessment was delayed to 2022.

Figure 3-1. Grade 4 and Grade 8 ELA Proficiency Rate: 2025 KAP vs. 2024 NAEP



Note. KAP = Kansas Assessment Program; NAEP = National Assessment of Educational Progress.

Figure 3-2. Grade 4 and Grade 8 Mathematics Proficiency Rate: 2025 KAP vs. 2024 NAEP



Note. KAP = Kansas Assessment Program; NAEP = National Assessment of Educational Progress.

3.3.3. Relationships Between KAP Assessment and KELPA

Relationships between KAP and KELPA performance can be used as another source of convergent evidence. In this section, the external assessments include KELPA grades 3–8 and 10. KELPA assesses EL students’ skills on ELA in four domains: listening, speaking, reading, and writing.

Pearson product-moment correlations were computed between KAP scale scores and KELPA domain scale scores in ELA, mathematics, and science. These correlations provide evidence of the relationship between English language proficiency and academic achievement.

Effect size interpretation (Cohen, 1988):

- Small: $r < 0.30$
- Medium: $0.30 \leq r < 0.50$
- Large: $r \geq 0.50$

Relationships were examined by domain because English learners’ proficiency in each domain in KELPA may differentially impact performance on grade-level academic assessments.

Table 3-3 presents correlation coefficients between KAP ELA scores and KELPA domain scores.

- The strongest correlations were observed between KAP ELA and the KELPA reading domain, ranging from 0.53 (grade 10) to 0.68 (grade 4).
- The weakest correlations were between KAP ELA and the speaking domain, ranging from 0.27 (grades 7 and 8) to 0.38 (grade 3).
- For listening and writing, medium to large correlation coefficients were observed across grades, indicating a consistent relationship between KAP ELA achievement and KELPA performance.
- Relationships between KAP ELA and KELPA were weakest overall at grade 10, with an average of 0.42 across domains.

Table 3-3. Correlations Between KAP ELA Scale Scores and KELPA Domain Scores by Grade

Grade	Listening	Speaking	Reading	Writing	Domain Average
3	0.48	0.38	0.66	0.61	0.53
4	0.54	0.34	0.68	0.61	0.54
5	0.52	0.32	0.66	0.57	0.52
6	0.59	0.31	0.65	0.54	0.52
7	0.56	0.27	0.67	0.52	0.50
8	0.46	0.27	0.56	0.43	0.43
10	0.44	0.28	0.53	0.42	0.42
Average	0.51	0.31	0.63	0.53	0.50

Table 3-4 presents correlations between KAP mathematics scores and KELPA domain scores.

- Compared to the relationships with KELPA ELA, relationships with KELPA mathematics were weaker across all domains.
- The strongest correlations were between KAP mathematics and the KELPA reading domain, ranging from 0.28 (grade 10) to 0.53 (grades 4 and 6).
- The weakest correlations were between KAP mathematics and the KELPA speaking domain, ranging from 0.15 (grade 10) to 0.33 (grade 3).
- Relationships between KAP mathematics and KELPA were weakest overall at grade 10, with an average of 0.24 across domains.

Table 3-4. Correlations Between KAP Mathematics Scale Scores and KELPA Domain Scores by Grade

Grade	Listening	Speaking	Reading	Writing	Domain Average
3	0.43	0.33	0.52	0.51	0.46
4	0.49	0.28	0.53	0.54	0.36
5	0.38	0.22	0.40	0.43	0.36
6	0.47	0.28	0.53	0.42	0.43
7	0.36	0.20	0.38	0.33	0.32
8	0.33	0.23	0.41	0.34	0.33
10	0.25	0.15	0.28	0.26	0.24
Average	0.41	0.26	0.46	0.43	0.39

Table 3-5 presents correlations between KAP science scores and KELPA domain scores.

- The strongest correlations were between KAP science and the KELPA reading domain.
- The weakest correlations were between KAP science and the KELPA speaking domain.

Table 3-5. Correlations Between KAP Science Scale Scores and KELPA Domain Scores by Grade

Grade	Listening	Speaking	Reading	Writing	Domain Average
5	0.50	0.29	0.53	0.47	0.45
8	0.46	0.27	0.54	0.39	0.41
Average	0.48	0.28	0.54	0.43	0.43

Across the four KELPA domains and grade levels, KAP ELA demonstrated the highest average correlation (0.50), followed by science (0.43), with mathematics showing the lowest (0.39). These values offer meaningful evidence of convergent validity, particularly in ELA, suggesting that KAP assessments are capturing constructs related to language proficiency and academic content. The stronger correlations in ELA are consistent with expectations, given the shared linguistic demands between English language proficiency assessments and ELA content.

3.4. Validity Evidence Based on Consequences of Testing

Validity evidence based on consequences refers to evidence supporting the intended uses and interpretation of test scores. A primary intended use of KAP test scores is to provide scores that can be used with local assessment scores to assist in improving a building’s or district’s programs as stated in the [KAP Test Coordinator Manual](#). Section 4.4. Scoring and Scaling summarizes how items and tests are scored. For a given test score, the performance level is determined by a set of established cut scores. Chapter 6 | Academic Achievement Standards and Reporting summarizes the cut scores and includes an example of a KAP student score report. To help educators and parents interpret KAP results, KAP also provides the [KAP Educator Guide](#) and the [KAP Parent Guide](#).

Chapter 4 | Technical Quality: Other

This chapter mainly describes evidence related to the technical quality of the Kansas Assessment Program (KAP) and summarizes the technical analysis. Most of the analysis described in this chapter is based on 2025 assessment data. Evidence for technical quality includes test reliability, fairness and accessibility, an item analysis summary, a test analysis summary, and trend data.

4.1. Reliability

Reliability is a test score consistency index that shows the degree of test score consistency across repeated measures. Test scores that are stable across repeated measures indicate a more reliable test. Factors leading to unstable test scores are called *measurement errors*. Measurement errors include, but are not limited to, changes in testing conditions; changes in a student’s knowledge, physical condition, or mental status; and changes in testing content across multiple test administrations.

Measurement errors cannot be fully removed but can be reduced. For example, standardized testing procedures reduce measurement errors caused by changing testing conditions. KAP has standardized testing procedures, and the same procedures are applied to all students; specific accommodations are provided to students with special needs. The [KAP Test Coordinator Manual](#) describes these testing procedure specifications.

In the context of educational achievement tests, factors such as learning, fatigue, and motivation may affect test takers at different rates for repeated measures. It is impractical to test the same content area repeatedly, as test takers cannot maintain the same knowledge, physical condition, and mental status across test administrations. Therefore, the reliability for educational measures is typically estimated rather than calculated directly. Estimated reliability coefficients range from 0 to 1. Higher values indicate more reliable tests with less measurement error.

In this section, we present reliability estimates for overall scores and subscores provided by the KAP assessments. The overall score reliability estimates are calculated for the full sample of tested students, as well as for student groups. We also include item response theory (IRT) information functions and conditional standard errors of measurement at each cut score, as well as classification consistency and accuracy estimates for overall scores. Finally, reliability, classification consistency, and accuracy estimates for KAP subscores are summarized.

4.1.1. Test Reliability

We used a marginal-reliability method (Green et al., 1984) to estimate test reliability. This method can estimate reliability for both fixed-form and adaptive tests. The calculation formula for marginal reliability is

$$\bar{\rho} = \frac{\sigma_{\theta}^2 - \overline{SE_{\theta}^2}}{\sigma_{\theta}^2}. \quad (\text{Equation 4-1})$$

The equation shows that marginal reliability, $\bar{\rho}$, is defined by two values: the variance of theta (σ_{θ}^2) and standard errors (SEs) of theta (SE_{θ}^2). Because standard errors are different across thetas, the mean of squared SEs, $\overline{SE_{\theta}^2}$, is used in the equation.

As shown in Table 4-1, marginal-reliability estimates for English language arts (ELA) ranged from 0.74 in grade 7 to 0.86 in grade 3. For mathematics, reliability estimates ranged from 0.78 at the high school level to 0.87 in grades 5 and 8. Science reliability estimates were 0.75 for grade 5 and 0.86 for grade 8.

Table 4-1. Test Reliability Estimate by Subject and Grade

Grade	English Language Arts	Mathematics	Science
3	0.86	0.84	—
4	0.82	0.85	—
5	0.82	0.87	0.75
6	0.80	0.84	—
7	0.74	0.83	—
8	0.78	0.87	0.86
10	0.75	0.78	—

Student-Group Reliability

We estimated reliabilities by the marginal-reliability method for gender groups, race groups, ethnicity groups, English learner (EL) status groups, and disability status groups.⁴ Table 4-2, Table 4-3, and Table 4-4 present student-group reliability estimates for ELA, mathematics, and science.

Across ELA, mathematics, and science, reliability coefficients were uniformly high, demonstrating strong internal consistency and dependable score precision across all grades and subgroups. Overall, reliability values ranged from approximately 0.69 to 0.89, with most falling between 0.80 and 0.86, indicating stable measurement characteristics. As in prior years, reliability was slightly higher in lower grades (3–5) and declined modestly

⁴ Economically disadvantaged status is not shared with AAI to protect the privacy of students, so this student group is not included in the comparison.

in upper grades (7–10).

Across subjects, mathematics tended to show the highest overall reliabilities, with values up to 0.89 for ELs in grade 3, and consistently above 0.80 for nearly all subgroups through grade 8. ELA reliability estimates were consistently strong, typically exceeding 0.80, with slight decreases observed in the upper grade levels. Science reliabilities were generally strong, ranging from approximately 0.80 and 0.88 across most subgroups. However, the lowest values were observed among Asian students in grade 7 (~0.69) and among ELs and students with disabilities in grade 10 (~0.72–0.74).

Table 4-2. Student-Group Reliability Estimates for English Language Arts

Subgroup	Grade						
	3	4	5	6	7	8	10
Gender							
Male	0.87	0.83	0.82	0.80	0.75	0.78	0.76
Female	0.86	0.82	0.82	0.79	0.74	0.77	0.75
Race							
NA	0.88	0.85	0.85	0.82	0.81	0.80	0.80
Asian	0.85	0.80	0.80	0.78	0.69	0.75	0.69
Black	0.88	0.85	0.85	0.82	0.80	0.80	0.79
NHPI	0.88	0.83	0.84	0.82	0.79	0.80	0.79
Other	0.87	0.83	0.83	0.81	0.76	0.78	0.76
White	0.86	0.82	0.82	0.80	0.74	0.77	0.75
Hispanic							
Yes	0.88	0.85	0.84	0.82	0.80	0.80	0.79
No	0.86	0.82	0.82	0.79	0.73	0.77	0.74
SWD							
Yes	0.88	0.85	0.85	0.82	0.82	0.80	0.80
No	0.86	0.82	0.82	0.79	0.73	0.77	0.75
EL							
Yes	0.89	0.86	0.86	0.83	0.84	0.81	0.81
No	0.86	0.82	0.82	0.79	0.73	0.77	0.75

Note. NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

Table 4-3. Student-Group Reliability Estimates for Mathematics

Subgroup	Grade						
	3	4	5	6	7	8	10
Gender							
Male	0.83	0.84	0.86	0.83	0.83	0.87	0.78
Female	0.85	0.87	0.87	0.85	0.84	0.88	0.79
Race							
NA	0.86	0.88	0.88	0.86	0.84	0.88	0.76
Asian	0.80	0.80	0.84	0.78	0.80	0.84	0.76
Black	0.86	0.88	0.88	0.87	0.84	0.88	0.75
NHPI	0.85	0.86	0.88	0.85	0.83	0.88	0.76
Other	0.84	0.86	0.87	0.85	0.84	0.87	0.78
White	0.84	0.85	0.87	0.84	0.83	0.87	0.79
Hispanic							
Yes	0.86	0.88	0.88	0.87	0.84	0.88	0.76
No	0.83	0.84	0.86	0.83	0.83	0.87	0.79
SWD							
Yes	0.85	0.87	0.86	0.86	0.83	0.87	0.72
No	0.83	0.85	0.87	0.84	0.83	0.87	0.79
EL							
Yes	0.86	0.88	0.88	0.87	0.84	0.88	0.72
No	0.83	0.85	0.87	0.83	0.83	0.87	0.79

Note. NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

Table 4-4. Student-Group Reliability Estimates for Science

Subgroup	Grade	
	5	8
Gender		
Male	0.75	0.86
Female	0.76	0.86
Race		
Native American	0.75	0.85
Asian	0.75	0.86
Black	0.74	0.85
NHPI	0.73	0.86
Other	0.75	0.86
White	0.76	0.86
Hispanic		
Yes	0.74	0.85
No	0.76	0.87
SWD		
Yes	0.73	0.84
No	0.76	0.87
EL		
Yes	0.73	0.83
No	0.76	0.87

Note. NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

4.1.2. Test Information

For KAP tests, we use IRT models to estimate students' latent ability levels, denoted as theta, which are then transformed to scale scores. These models also allow for the calculation of test information functions (TIFs), which quantify the precision of measurement across the full performance continuum. A TIF is derived by summing the item information functions of all operational items within a grade-level test.

The TIF indicates how much information the test provides at each theta value and serves as an IRT-based analogue to the reliability coefficient in classical test theory. Figure 4-1, Figure 4-2, and Figure 4-3 present the TIFs for theta values ranging from -3 to 3 in increments of 0.5 for each grade in ELA, mathematics, and science. These figures also mark the level 3 theta thresholds, which correspond to proficiency cut scores.

Typically, TIF values peak near the center of the theta distribution and taper off toward the extremes, producing a bell-shaped curve. Across all subjects and grades, the tests exhibit well-defined information curves centered around ability levels most relevant for differentiating student performance.

ELA grades 5, 7, and 10 demonstrate the highest information levels, with all grades showing broad coverage across the theta continuum. Lower ELA grades maintain consistent precision near average ability levels, supporting reliable measurement across developmental stages.

In mathematics, most grades show peak information between theta values of -0.5 and -1 . Grades 3–5 yield the highest information overall. Grade 10 mathematics is distinct in that its information function peaks at a theta of approximately 0.5 . Despite this variation, all mathematics grades maintain similar patterns, offering steady information near the proficiency range.

For science, grades 5 and 8 both reach peak information at $\theta \approx 0$, providing optimal precision at the ability level where most students are typically located.

Figure 4-1. Test Information Function for English Language Arts

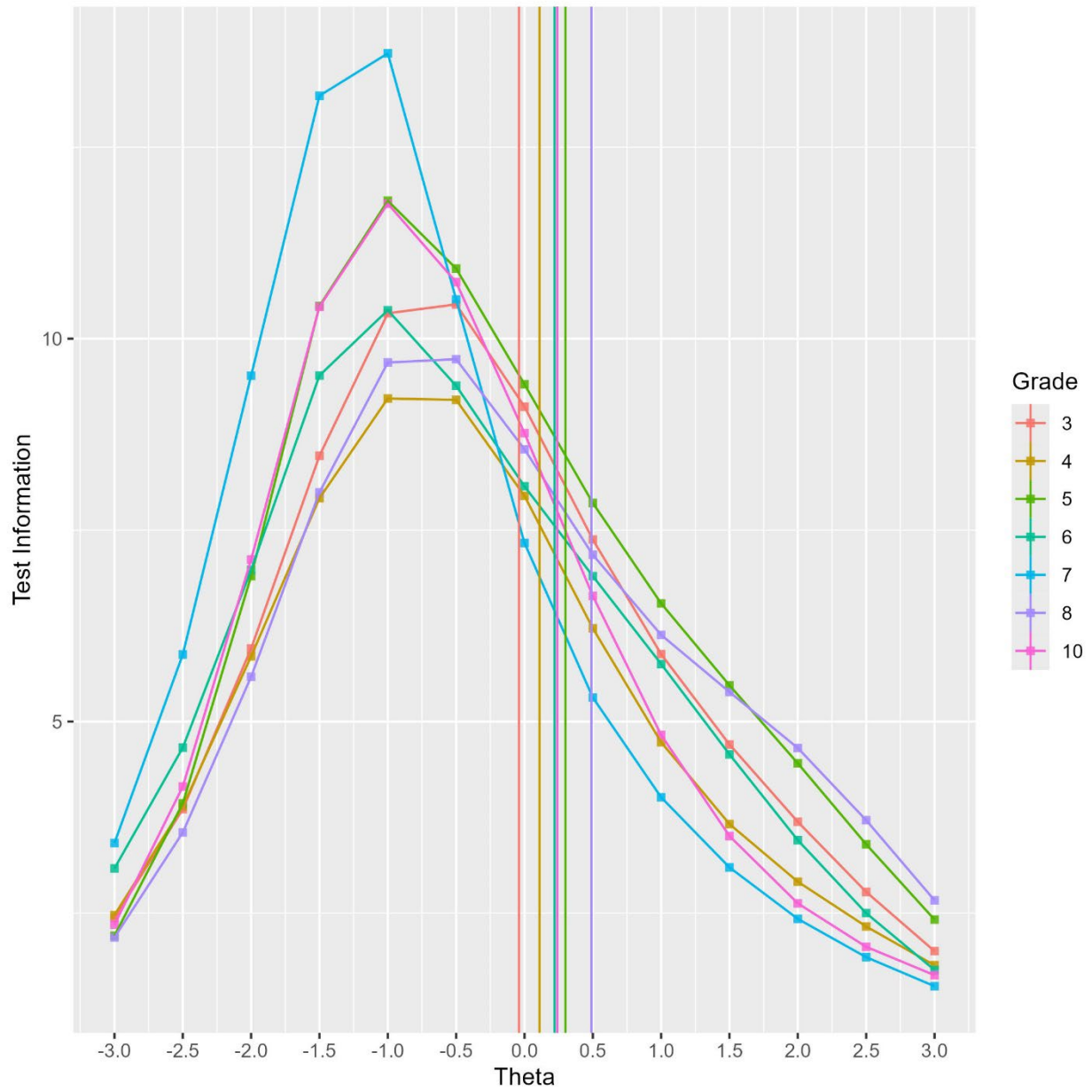


Figure 4-2. Test Information Function for Mathematics

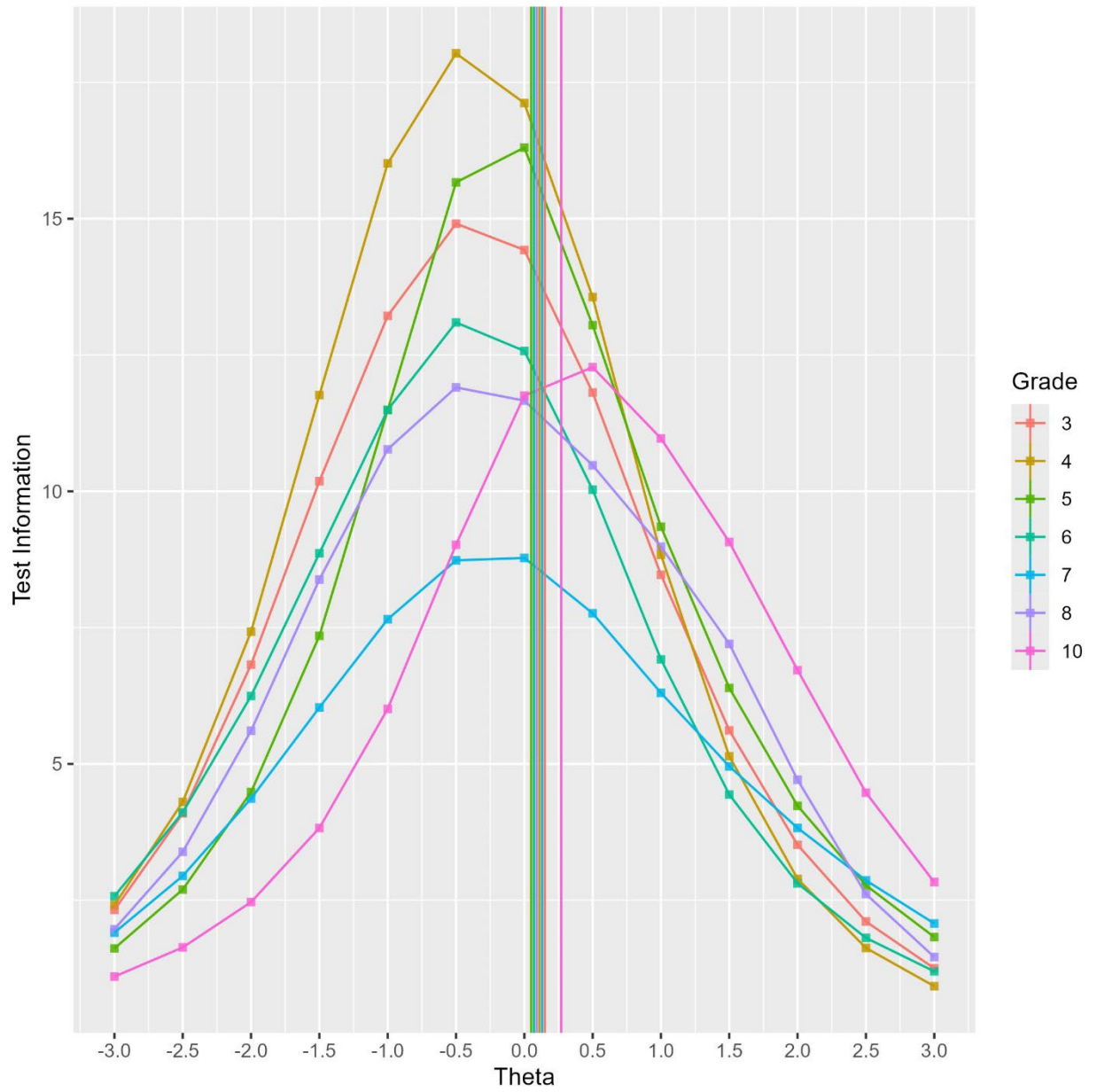
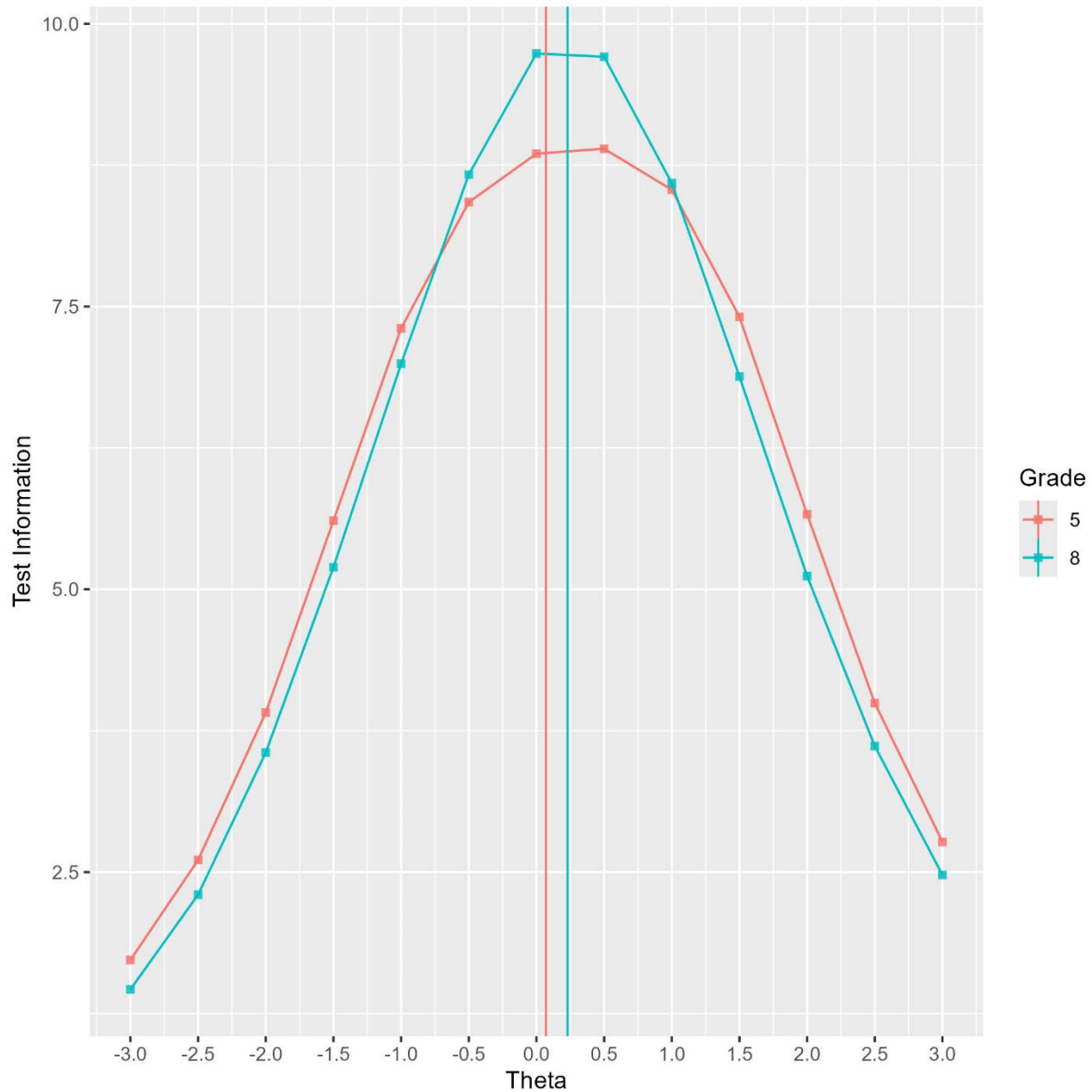


Figure 4-3. Test Information Function for Science



In IRT, the conditional standard error of measurement (CSEM) is estimated for each value of theta, reflecting the precision of ability estimates at different performance levels. CSEMs are inversely related to TIFs; as TIF increases, CSEM decreases. For reporting purposes, CSEMs are converted to the scale-score metric, as shown in Table 4-5.

Across all subjects, CSEMs tend to increase with performance level—from level 2 to 4—indicating that measurement precision generally declines at higher cut scores. Among the three subjects, mathematics exhibits the lowest CSEMs, while ELA shows the highest. Within each subject, variation in CSEM is more pronounced across performance levels than across grade levels.

Table 4-5. CSEM Across Subjects, Performance Levels, and Grade Levels

Grade	English Language Arts			Mathematics			Science		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	15.3	16.4	23.4	13.2	13.4	17.2	—	—	—
4	16.3	18.2	28.0	12.0	12.3	19.0	—	—	—
5	16.8	18.6	23.7	13.4	12.5	22.2	18.0	16.9	17.8
6	15.8	18.2	24.9	14.3	14.2	23.8	—	—	—
7	13.9	20.0	29.1	17.4	17.1	23.8	—	—	—
8	17.7	19.6	24.8	14.8	14.7	19.7	17.2	15.9	16.7
10	14.8	18.0	29.1	17.2	14.2	19.3	—	—	—

4.1.3. Classification Consistency and Accuracy

Classification consistency and accuracy indicate how accurately students are classified into performance levels. Performance-level classification consistency and accuracy are of great interest for testing programs that serve as accountability purposes. According to Livingston and Lewis (1995), *classification consistency* refers to “the agreement between the classifications based on two nonoverlapping, equally difficult forms of the test” (p. 180), and *classification accuracy* refers to “the extent to which the actual classifications of test takers on the basis of their single-form scores agree with those that would be made on the basis of their true scores, if their true scores could somehow be known” (p. 180). Both classification consistency and accuracy indices range from 0 to 1, with 0 representing classifications that are not consistent or accurate and 1 representing perfectly consistent or accurate classifications.

Because of the unobservable nature of true scores and the impossibility of repeated testing, actual observed-score distribution and reliabilities are used to estimate a true-score distribution and an observed-score distribution for an alternate parallel form (Livingston & Lewis, 1995). Classification consistency is calculated as the classification agreement between two observed-score distributions (i.e., the observed-score distributions of actual and alternate parallel forms). Kappa is used to calculate the degree of agreement. Classification accuracy is calculated as the probability of accurate classification between the true score and actual observed-score distributions.

Table 4-6 presents the results for overall classification consistency and accuracy across all four performance levels, as well as for the dichotomies defined by the three cut scores. Across all subjects, accuracy (A) values remain consistently high—generally ranging from 0.63 to 0.97—demonstrating strong agreement between observed and model-predicted performance-level classifications. Consistency (C) values, which reflect the stability of classifications under repeated administrations, are lower and show more variation across grades and subjects. For ELA, overall consistency values range from 0.52 to 0.63, while values for mathematics range from 0.54 to 0.59 and science from 0.53 to 0.59. Across

subjects, accuracy is highest for the upper cut score (levels 1–3 vs. 4), whereas consistency increases slightly as cut-score groupings become broader. Overall, accuracy is uniformly high across grades and subjects, while consistency varies modestly, with math and science showing comparable or slightly greater stability than ELA.

Table 4-6. Overall Classification Consistency (C) and Accuracy (A)

Subject and Grade	Cut-Score Category							
	Overall		1 vs. 2, 3, 4		1, 2 vs. 3, 4		1, 2, 3 vs. 4	
	C	A	C	A	C	A	C	A
ELA								
3	0.63	0.72	0.84	0.88	0.84	0.88	0.93	0.95
4	0.59	0.69	0.81	0.86	0.82	0.87	0.92	0.95
5	0.61	0.70	0.83	0.88	0.81	0.87	0.92	0.94
6	0.57	0.68	0.82	0.87	0.81	0.86	0.92	0.95
7	0.54	0.64	0.78	0.84	0.78	0.84	0.93	0.95
8	0.56	0.66	0.79	0.85	0.80	0.86	0.93	0.95
10	0.52	0.63	0.76	0.82	0.80	0.86	0.93	0.95
Mathematics								
3	0.57	0.67	0.79	0.85	0.84	0.88	0.90	0.93
4	0.54	0.64	0.73	0.79	0.84	0.89	0.94	0.96
5	0.59	0.69	0.76	0.82	0.85	0.90	0.95	0.97
6	0.55	0.64	0.73	0.79	0.84	0.89	0.95	0.97
7	0.54	0.63	0.72	0.77	0.85	0.89	0.96	0.97
8	0.57	0.67	0.75	0.81	0.87	0.91	0.95	0.97
10	0.55	0.66	0.74	0.80	0.83	0.88	0.96	0.97
Science								
5	0.53	0.64	0.81	0.87	0.78	0.84	0.90	0.93
8	0.59	0.68	0.84	0.89	0.84	0.89	0.88	0.92

Note. ELA = English language arts.

4.1.4. Subscore Reliability

In addition to the total test score, the scores of subsets of ELA, mathematics, and science items are reported as subscores. The number of items in each subscore varies, and some items contribute to multiple subscores. The minimum number of items required to report a subscore is six.

ELA has eight subscores in total. Six are consistent across all grades (grades 3–8 and 10). The primary subscores are Overall Reading and Overall Writing. Overall Reading is further divided into Reading Literature and Reading Information, while Overall Writing includes Writing Text Types and Purposes.

Three cluster-level subscores—Key Ideas and Details, Craft and Structure, and Language in Reading—span both Overall Writing and Overall Reading. Table 4-7 shows subscores for each grade within the skills-and-concepts subscore for ELA.

Table 4-7. Subscores for English Language Arts by Grade

Subscore	Grade						
	3	4	5	6	7	8	10
Overall Writing	X	X	X	X	X	X	X
Writing Text Types and Purposes	X	X	X			X	X
Overall Reading	X	X	X	X	X	X	X
Reading Literature	X	X	X	X	X	X	X
Reading Information	X	X	X	X	X	X	X
Key Ideas & Details	X	X	X	X	X	X	X
Craft & Structure	X	X	X	X	X	X	X
Language in Reading	X	X	X				
Number of subscores	8	8	8	6	6	7	7

The number of mathematics subscores varies by grade. Each of grades 3, 7, 8, and 10 include six subscores, while grades 4–6 include seven. Table 4-8 shows subscores for each grade within the skills-and-concepts subscore for mathematics.

Table 4-8. Subscores for Mathematics by Grade

Subscore	Grade						
	3	4	5	6	7	8	10
Operations & Algebraic Thinking	X	X					
Operations & Algebraic Thinking: Multiplication & Division	X						
Operations & Algebraic Thinking: Operations with Whole Numbers		X					
Number & Operations in Base 10	X	X	X				
Number & Operations: Using Place Value in Arithmetic		X					
Number & Operations: Multi-digit Numbers & Decimals			X				
Number & Operations - Fractions	X	X	X				
Number & Operations - Fractions: Building & Operating on Fractions		X					
Number & Operations - Fractions: Decimal Notation		X					
Number R Operations: Adding & Subtracting Fractions			X				
Number & Operations: Multiplying & Dividing Fractions			X				
Measurement & Data	X		X				
Measurement & Data: Area	X						
Measurement & Data: Volume			X				
Ratios & Proportional Relationships				X	X		
The Number System				X	X		
The Number System: Common Factors & Multiples				X			
The Number System: Rational Numbers				X			
Geometry					X	X	X
Geometry: Understanding Angles & Angle Measurements						X	
Expressions & Equations				X	X	X	
Expressions & Equations: Equations & Inequalities				X			
Expressions & Equations: Proportions & Linear Equations						X	
Functions						X	X
Functions: Interpreting Functions							X
Statistics & Probability				X	X		X
Statistics & Probability: Probability Models					X		
Statistics & Probability: Bivariate Data						X	
Algebra							X
Algebra: Reasoning with Equations & Inequalities							X
Number of subscores	6	7	7	7	6	6	6

Each science test includes seven subscores. organized under the major domains of Physical Science, Life Science, and Earth & Space Science. Each domain contains two designated subcategories that contribute to the overall subscore structure. Table 4-9 shows subscores for each grade within the skills-and-concepts subscore for science.

Table 4-9. Subscores for Science by Grade

Subscore	Grade	
	5	8
Earth & Space Science	X	X
Developing & Using Models		X
Using Evidence for Explanations & Argumentation	X	
Life Science	X	X
Developing & Using Models	X	
Using Evidence for Explanations & Argumentation		X
Physical Science	X	X
Developing & Using Models	X	X
Engaging in Inquiry	X	X
Number of subscores	7	7

Subscores are classified into three categories—below, near, and above—that are computed from the mean and standard error of measurement (SEM) of the subscores. The below cut point is defined as the ceiling of the subscore mean for the population, while the above cut point is defined as the ceiling of the mean plus two times the SEM.

Reliability analyses based on raw scores were conducted to evaluate the consistency of subscores. Coefficient alpha was used when a subscore included polytomously scored items, and the Kuder–Richardson Formula 20 (KR-20) was used when all items were dichotomously scored. Both indices estimate internal consistency. Internal consistency is evidence of the extent to which items within a subscore measure the same construct and produce consistent results. The KR-20 coefficient is derived from the proportion of students who answered each item correctly and the variance of total raw scores. Higher values of alpha or KR-20 indicate greater reliability of the subscore.

Across all subjects, reliabilities based on raw scores were generally moderate to high, indicating consistent measurement within subscores. ELA reliabilities ranged from 0.49 to 0.89, with the highest values observed for Overall Reading (≥ 0.85 across grades) and lower values for finer-grained writing and cluster scores. Mathematics reliabilities ranged from 0.40 to 0.84, with strong performance in domains such as Number and Operations and Geometry, and lower estimates for narrower subscores (e.g., Probability Models and Equations in middle grades). Science reliabilities ranged from 0.368 to 0.752, with higher values for broad domains (Physical Science and Earth & Space Science) and lower reliabilities for detailed subdomains. Appendix G | Subscore Reliability presents the results of the subscore reliability analysis.

Overall, reliabilities tended to increase with grade level and were consistently higher for total or domain-level scores than for subdomains, reflecting greater score stability when more items contributed to the subscore.

4.2. Accessibility and Fairness

During the development and administration of the KAP assessment, we considered accessibility for all students and fairness across student groups in every step. We used universal design (UD) as a guide during the development of items, test formats, and the online test delivery interface to ensure fairness and accessibility for all students. Section 4.2.2. Fairness summarizes the UD guidelines. All operational items pass a bias and sensitivity review to mitigate the likelihood of content bias toward any one student group. The bias and sensitivity review described in Section 2.2.3. Item Review has external reviewers review items to identify unfairness barriers that may prevent students from demonstrating what they know and can do.

4.2.1. Accessibility

According to the *Standards for Educational and Psychological Testing* (hereafter the *Standards*), “accessibility is the degree to which the items or tasks on a test enable as many test takers as possible to demonstrate their standing on the target construct without being impeded by characteristics of the item that are irrelevant to the construct being measured” (American Psychological Association [APA], 2014, p. 215). Evidence in support of accessibility of an assessment comprises inclusion, accommodations, and the implementation of UD in items and test development. *UD* refers to principles that provide equal access to all students. Section 4.2.2. Fairness summarizes the implementation of UD in item and test development. However, some barriers, such as blindness, cannot be addressed by UD. Test inclusion and accommodation policies help address these needs. The Kite® online test system provides many accommodations, including magnification, text-to-speech, and color contrast, among others. Some students require braille tests, which are made available to students who need them. For more details about accommodations for KAP, see Chapter 5 | Inclusion of All Students.

4.2.2. Fairness

According to the *Standards*, “the central idea of fairness in testing is to identify and remove construct-irrelevant barriers to maximal performance for any examinee” (APA et al., 2014, p. 74). The *Standards* identifies fairness as an issue related to the validity of test-score inferences. Evidence supporting the assertion of fairness in an assessment comes from several stages, such as the item and test development stages before test administration and the differential item functioning (DIF) analysis stage after test administration.

Using appropriate item and test development processes is an excellent start for ensuring fairness. UD in item and test development not only allows for the participation of the widest range of students, but also bolsters the validity of score inferences. KAP’s comprehensive inclusion rules mean that KAP tests include all Kansas students (details about the policy of including all students are in Chapter 5 | Inclusion of All Students). While the initial intention is to meet the assessment requirements of special-needs students, the benefits of universally designed assessments should apply to all students with diverse characteristics. Item writer training informs participants about UD concepts, includes a

definition of UD, and provides examples of test items that adhere to UD principles. Additionally, item writer guidelines comprise many UD principles. The following are UD guidelines used during KAP test development:

- Item writers are trained to become aware of, and sensitive to, issues of cultural and regional diversity.
- Both internal and external reviewers of items and test specifications strive to ensure that no barriers stem from a lack of sensitivity to ability, culture, or other characteristics.
- The tests are compatible with many accommodations and a variety of widely used adaptive equipment and assistive technology without changing the meaning or difficulty of test items.
- The language used in test materials is direct and concise. Additionally, unnecessary images and text are omitted to avoid distracting students.

For DIF results, see Section 3.2.3. Differential Item Functioning. DIF analyses conducted for the current administration indicate that no items were identified with significant DIF across gender (i.e., female vs. male), race (i.e. Black vs. White), and EL status (i.e., EL vs. non-EL) for all three subjects. DIF analysis examines whether an item shows any statistical difference between two groups of students after controlling for student proficiency. No items with DIF contribute to the evidence in support of fairness during item writing and reviewing.

4.3. Full Performance Continuum

KAP was designed and developed to produce a reasonably precise estimation of student proficiency across the full performance continuum in each subject area and grade. TIFs across different ability levels and conditional errors of measurement at the cut scores from Section 4.1.2. Test Information show test precision across the full range of ability estimates. Results indicate that KAP assessments can accurately estimate ability across the full theta scale, especially in the middle of the scale.

During test construction, we screen item quality through item difficulty, item-total correlation, DIF, option analyses, and IRT parameters. This approach not only ensures the quality of items to be used on the test but also provides the widest range possible for measuring student abilities. Additionally, we plot test characteristic curves, test information, and CSEM during test construction to gauge the proficiency range of each test. To confirm that the tests efficiently cover the full performance continuum as expected, we present classical and IRT item statistics as evidence.

4.3.1. Classical Item Statistics

Here we calculate and provide two statistics: item difficulty and item discrimination. *Item difficulty* refers to the difficulty of an item, and *item discrimination* indicates the degree to which an item differentiates between students with high proficiency and those with low

proficiency. Item difficulty in classical test theory is expressed as a p value or mean score. A p -value is the percentage of students who answer the item correctly. Equation 4-2 shows the calculation of the p -value.

$$p\text{-value} = \frac{\frac{1}{n} \sum_{i=1}^n x_i}{\text{item max score}}, \quad (\text{Equation 4-2})$$

where x refers to the observed score,
 i refers to student i , and
 n refers to the total number of students who took the item.

Table 4-10, Table 4-11, and Table 4-12 present summaries of item difficulty for ELA, mathematics, and science tests. The mean item difficulties per grade range from 0.51 (grade 10) to 0.58 (grade 7) for ELA, from 0.36 (grade 10) to 0.58 (grade 4) for mathematics, and from 0.36 (grade 8) to 0.41 (grade 5) for science. For all grades and subjects, the ranges of item difficulty are large, ranging from 0.05 to 0.88 for ELA, from 0.06 to 0.88 for mathematics, and from 0.08 to 0.76 for science.

Table 4-10. Summary Statistics for Classical Item Difficulties for ELA

Grade	No. of Items	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	40	0.53	0.15	0.14	0.44	0.55	0.65	0.78
4	40	0.55	0.16	0.09	0.49	0.56	0.64	0.86
5	40	0.56	0.17	0.16	0.46	0.56	0.67	0.88
6	40	0.55	0.16	0.05	0.47	0.54	0.65	0.85
7	40	0.58	0.20	0.06	0.43	0.59	0.75	0.87
8	40	0.54	0.14	0.26	0.44	0.54	0.65	0.79
10	40	0.51	0.21	0.06	0.37	0.54	0.64	0.87

Note. M = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-11. Summary Statistics for Classical Item Difficulties for Mathematics

Grade	No. of Items	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	40	0.56	0.16	0.17	0.46	0.56	0.67	0.85
4	40	0.58	0.14	0.34	0.46	0.58	0.65	0.88
5	40	0.47	0.17	0.19	0.33	0.46	0.58	0.84
6	40	0.54	0.15	0.22	0.41	0.54	0.63	0.86
7	40	0.44	0.17	0.06	0.34	0.45	0.52	0.82
8	40	0.49	0.16	0.11	0.38	0.50	0.61	0.83
10	40	0.36	0.14	0.09	0.26	0.32	0.48	0.61

Note. M = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-12. Summary Statistics for Classical Item Difficulties for Science

Grade	No. of Items	<i>M</i>	<i>SD</i>	Min	<i>P</i> ₂₅	Median	<i>P</i> ₇₅	Max
5	40	0.41	0.21	0.09	0.23	0.39	0.58	0.76
8	40	0.36	0.17	0.08	0.27	0.33	0.50	0.76

Note. *M* = mean; *SD* = standard deviation; *P*₂₅ = 25th percentile; *P*₇₅ = 75th percentile.

Item discrimination reflects an item’s ability to differentiate students of high proficiency from those of low proficiency. Ideally, high-achieving students (i.e., those with high raw scores) should be more likely to answer any given item correctly, whereas low-achieving students (i.e., those with low raw scores) should be more likely to answer the same item incorrectly. The Pearson’s product-moment correlation coefficient between student item scores and test scores is also referred to as item-total correlations, although strictly speaking these are point-biserial correlations when items have dichotomous (0, 1) scores.

The item-total correlation is used as an index of item discrimination. The item-total correlation ranges from –1.0 to 1.0. Positive values indicate that students with higher raw scores are more likely to answer an item correctly than those with lower raw scores; negative values indicate the opposite. The magnitude of the correlation indicates the degree of discrimination, in that items with higher values have better discrimination power. The information on measuring the full performance continuum is not directly provided by classical test theory (CTT) item discrimination, but a test with more high-discrimination items will provide more accurate measures of proficiency than a test with lower-discriminating items.

Table 4-13, Table 4-14, and Table 4-15 present item discrimination for the three subjects. The mean item discrimination across grades ranges from 0.41 (grade 4) to 0.46 (grades 7 and 8) for ELA, from 0.38 (grade 7) to 0.50 (grade 4) for mathematics, and from 0.38 (grade 8) to 0.40 (grade 5) for science. The minimum item discrimination is 0.08 at grade 5 science, and the maximum is 0.67 at grade 8 ELA and grade 5 math.

Table 4-13. Summary Statistics for Classical Item Discrimination for ELA

Grade	No. of Items	<i>M</i>	<i>SD</i>	Min	<i>P</i> ₂₅	Median	<i>P</i> ₇₅	Max
3	40	0.43	0.11	0.17	0.36	0.46	0.50	0.62
4	40	0.41	0.10	0.18	0.34	0.42	0.49	0.61
5	40	0.45	0.10	0.17	0.41	0.46	0.52	0.63
6	40	0.44	0.09	0.20	0.38	0.43	0.50	0.63
7	40	0.46	0.09	0.20	0.40	0.47	0.51	0.66
8	40	0.46	0.10	0.23	0.40	0.47	0.54	0.67
10	40	0.42	0.11	0.18	0.34	0.45	0.51	0.60

Note. *M* = mean; *SD* = standard deviation; *P*₂₅ = 25th percentile; *P*₇₅ = 75th percentile.

Table 4-14. Summary Statistics for Classical Item Discrimination for Mathematics

Grade	No. of Items	<i>M</i>	<i>SD</i>	Min	<i>P</i> ₂₅	Median	<i>P</i> ₇₅	Max
3	40	0.48	0.09	0.30	0.44	0.49	0.53	0.66
4	40	0.50	0.09	0.29	0.45	0.50	0.56	0.69
5	40	0.47	0.12	0.17	0.39	0.47	0.56	0.67
6	40	0.44	0.11	0.19	0.38	0.45	0.51	0.66
7	40	0.38	0.12	0.12	0.30	0.40	0.49	0.57
8	40	0.44	0.11	0.21	0.37	0.45	0.52	0.63
10	40	0.44	0.12	0.19	0.36	0.45	0.53	0.66

Note. *M* = mean; *SD* = standard deviation; *P*₂₅ = 25th percentile; *P*₇₅ = 75th percentile.

Table 4-15. Summary Statistics for Classical Item Discrimination for Science

Grade	No. of Items	<i>M</i>	<i>SD</i>	Min	<i>P</i> ₂₅	Median	<i>P</i> ₇₅	Max
5	40	0.40	0.12	0.08	0.32	0.43	0.50	0.56
8	40	0.38	0.12	0.09	0.32	0.38	0.47	0.58

Note. *M* = mean; *SD* = standard deviation; *P*₂₅ = 25th percentile; *P*₇₅ = 75th percentile.

4.3.2. Item Response Theory Item Statistics

KAP uses the two-parameter logistic (2PL) IRT model and its polytomous counterpart, the graded response model, as measurement models. For those two IRT models, item parameters include item difficulty (i.e., *b* parameter) and item discrimination (i.e., *a* parameter). Section 3.2.2. Item Response Theory and Model Assumptions has more detailed information about these two IRT models and their item parameters.

Table 4-16, Table 4-17, and Table 4-18 summarize the difficulty (i.e., *b* parameter) estimates of operational items in ELA, mathematics, and science tests, respectively. The IRT *b* parameter ranges from negative infinity to positive infinity and is on the same scale as the ability estimates. The higher the *b* parameter, the more difficult the item. Most items are dichotomous, but some items have as many as 11 score categories (thus, 10 *b* parameters yet still only one *a* parameter); therefore, the numbers of *b* and *a* parameters are different in these tables. Parameters for all items, irrespective of the number of score categories, are included together in Table 4-16, Table 4-17, and Table 4-18.

Mean IRT item difficulties across grades range from -0.61 to -0.32 for ELA, from -0.41 to 0.23 for mathematics, and from -0.33 to 0.04 for science. For all grades and subjects, the ranges of item difficulties are large, ranging from -6.78 to 4.88 for ELA, from -7.32 to 3.33 for mathematics, and from -5.10 to 3.41 for science. The large IRT item difficulty ranges indicate that the items included in KAP assessments adequately cover the full performance continuum.

Table 4-16. Summary Statistics for Item Response Theory Item Difficulty for ELA

Grade	No. of <i>b</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	45	0.10	1.12	-2.38	-0.68	-0.19	0.63	3.25
4	42	-0.21	1.14	-3.19	-0.86	-0.35	0.18	3.18
5	41	-0.16	1.04	-1.55	-0.82	-0.29	0.14	4.37
6	42	-0.36	0.99	-2.65	-0.96	-0.28	0.10	3.40
7	43	-0.41	1.41	-4.77	-1.30	-0.51	0.34	3.31
8	40	-0.27	0.65	-1.54	-0.74	-0.30	0.26	0.95
10	43	-0.01	1.33	-2.62	-0.88	-0.26	0.53	4.41

Note. *b* = difficulty parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-17. Summary Statistics for Item Response Theory Item Difficulty for Mathematics

Grade	No. of <i>b</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	40	-0.34	0.87	-3.08	-0.81	-0.28	0.16	1.56
4	40	-0.40	0.64	-2.42	-0.71	-0.43	0.12	0.59
5	40	0.24	1.02	-1.94	-0.41	0.12	1.12	2.63
6	40	-0.21	0.87	-1.85	-0.64	-0.32	0.23	2.46
7	42	0.63	1.83	-1.82	-0.26	0.19	1.17	7.13
8	40	0.04	1.01	-1.42	-0.64	-0.13	0.43	4.08
10	41	0.91	1.68	-2.00	0.02	0.82	1.17	9.21

Note. *b* = difficulty parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-18. Summary Statistics for Item Response Theory Item Difficulty for Science

Grade	No. of <i>b</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
5	47	0.77	1.59	-2.40	-0.38	0.64	1.68	4.02
8	42	0.38	0.88	-1.35	-0.23	0.38	0.81	2.82

Note. *b* = difficulty parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-19, Table 4-20, and Table 4-21 summarize the discrimination (i.e., *a* parameter) estimates of items in ELA, mathematics, and science tests, respectively. The IRT *a* parameter reflects an item's ability to differentiate students of high ability from those of low ability. Higher values indicate better discrimination power.

As with CTT item discrimination, the information measuring the full performance continuum is not directly provided by IRT *a* parameters, but a test with more items having high item discrimination will provide more accurate measures of proficiency than a test with fewer discriminating items.

The mean IRT item discrimination values range from 0.99 (grade 3) to 1.22 (grade 7) for ELA, from 0.94 (grade 7) to 1.43 (grade 4) for mathematics, and from 1.00 (grade 5) to 1.02 (grade 8) for science. Across all subjects and grades, the minimum discrimination parameters typically fall between 0.20 and 0.60. Notable outliers include grade 7 mathematics (0.12) and grade 4 mathematics (0.69). Although mean discrimination values cluster near 1.0, the parameter exhibits meaningful variability across items, supporting the use of the 2PL model, which allows discrimination to vary by item.

Table 4-19. Summary Statistics for Item Response Theory Item Discrimination for ELA

Grade	No. of <i>a</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	44	0.99	0.45	0.24	0.62	1.03	1.30	1.90
4	40	1.02	0.44	0.25	0.70	0.94	1.37	1.96
5	40	1.20	0.48	0.33	0.92	1.16	1.41	2.59
6	40	1.09	0.43	0.45	0.80	0.98	1.28	2.41
7	40	1.22	0.49	0.51	0.89	1.15	1.37	2.78
8	40	1.16	0.41	0.39	0.85	1.16	1.40	2.21
10	40	1.14	0.48	0.43	0.73	1.08	1.37	2.31

Note. *a* = discrimination parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-20. Summary Statistics for Item Response Theory Item Discrimination for Mathematics

Grade	No. of <i>a</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
3	40	1.30	0.41	0.25	1.10	1.26	1.57	2.20
4	40	1.43	0.44	0.69	1.17	1.41	1.68	2.58
5	40	1.30	0.55	0.27	0.92	1.23	1.64	2.70
6	40	1.16	0.43	0.33	0.91	1.14	1.44	2.16
7	40	0.94	0.44	0.12	0.59	0.98	1.31	1.70
8	40	1.19	0.53	0.25	0.74	1.20	1.56	2.36
10	40	1.11	0.54	0.25	0.73	1.03	1.49	2.46

Note. *a* = discrimination parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

Table 4-21. Summary Statistics for Item Response Theory Item Discrimination for Science

Grade	No. of <i>a</i> Parameters	<i>M</i>	SD	Min	P ₂₅	Median	P ₇₅	Max
5	42	1.00	0.46	0.22	0.62	0.93	1.39	1.91
8	40	1.02	0.34	0.32	0.84	1.01	1.21	1.76

Note. *a* = discrimination parameter; *M* = mean; SD = standard deviation; P₂₅ = 25th percentile; P₇₅ = 75th percentile.

4.4. Scoring and Scaling

This section introduces the procedures of scoring individual items, scoring the whole test, and scaling. We include the test results and performance-level distribution for 2025 KAP testing and present the KAP performance trend for the previous five years. Finally, this section describes the quality-control procedures used to ensure the accuracy of scoring and scaling.

4.4.1. Scoring

Item and test scoring in the 2025 administration remained the same as in previous years. All items were machine scored. The same test scoring method used previously was used this year.

Item Scoring

All KAP assessment items administered in 2025 were machine scored. The online test delivery platform compared student responses to the correct keys stored with the items and assigned the scores accordingly.

Test Scoring

Test scoring used a psychometric model to derive item scores on the test to produce a single score indicating a student's proficiency level. We computed the IRT ability estimates (i.e., thetas) using the 2PL model and GRM. Because the total score was derived using the summed-score method (Thissen & Wainer, 2001)—in which scores for each item were added together to derive the raw score—thetas had a one-to-one correspondence with raw scores (i.e., each raw score has only one matching theta). By using the test-characteristic curve function of the IRT models, we obtained the theta for each raw score point for a test form (Press et al., 2007).

4.4.2. Scaling

Scaling is the process of transforming thetas (θ) to a reporting scale to facilitate score use and interpretation. The purpose of scaling is to express student performance on a scale that is intuitive and comparable across years.

Theoretical theta values range from negative to positive infinity and may include negative values and decimals, which can be difficult to interpret. To support score interpretation, theta values are linearly transformed into a reporting scale composed of positive integers. To maintain separation from the previous KAP summative scale, the new scale has a lowest obtainable scale score (LOSS) of 400 and a highest obtainable scale score (HOSS) of 700.

The distance between level 2 and level 3 cuts was fixed at 30 scale-score points. The new 2025 scale was therefore anchored at both the level 2 and level 3 cuts so that score gains could be expressed in terms of the distance between these two key performance levels.

The following sections describe the process used to construct the 2025 KAP scale scores.

Scale Transformation

A linear transformation was applied to convert theta (θ) values to the reporting scale:

$$SS = A \times \theta + B \quad \text{(Equation 4-3)}$$

where SS = scale score,

θ = estimated IRT ability (theta),

A = slope (scale-transformation constant), and

B = intercept (scale-transformation constant).

Rather than using standard deviations to compute the scaling constants, the Two Known Points method was used. In this method, two known points—typically thetas corresponding to the level 2 and level 3 cuts established in standard setting—are linked to their target scale-score values.

If (θ_1, SS_1) represents the level 2 cut (510) and (θ_2, SS_2) represents the Level 3 cut (540), then the slope and intercept are determined as follows:

$$A = \frac{SS_2 - SS_1}{\theta_2 - \theta_1} \quad \text{(Equation 4-4)}$$

$$B = SS_1 - A \times \theta_1 \quad \text{(Equation 4-5)}$$

Constants A and B ensure that the transformation passes exactly through the two anchor points defined by the level 2 and level 3 cuts. Once these constants are obtained, all theta values are transformed to the reporting scale using Equation 4-3.

Scale-Transformation Constant

The scoring process described in the Test Scoring section produces theta estimates for all students. Thetas associated with level 2 and level 3 cuts were identified for each grade and subject during the 2025 standard-setting study (see Table 4-22, Table 4-23, and Table 4-24). These two points served as anchors for constructing the linear transformation.

For each grade and subject, the slope A and intercept B were computed using equations 4-4 and 4-5. The resulting constants define the relationship between theta and the 2025 KAP scale score.

Once A and B are known, other cut scores (for example, the level 1/2 and level 3/4 boundaries) can be obtained by substituting the corresponding theta values into Equation 4-3. Because the resulting scale-score cuts may include decimal fractions, operational cut scores were rounded up to the next integer. This rounding convention ensures that students must earn a score equal to or greater than the cut to reach the higher performance level.

Table 4-25 presents the slope (*A*) and intercept (*B*) for all grades and subjects used to generate the 2025 KAP scale scores.

Table 4-22. English Language Arts Cut Scores

Grade	Theta Cuts			Scale-Score Cuts		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	-0.74	-0.04	1.56	510	540	609
4	-0.65	0.11	1.80	510	540	607
5	-0.54	0.30	2.01	510	540	600
6	-0.77	0.22	1.74	510	540	593
7	-0.83	0.24	2.15	510	540	600
8	-0.65	0.49	2.32	510	540	594
10	-0.82	0.24	1.79	510	540	588

Table 4-23. Mathematics Cut Scores

Grade	Theta Cuts			Scale-Score Cuts		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	-0.71	0.15	1.00	510	540	570
4	-0.75	0.11	1.23	510	540	579
5	-0.73	0.05	1.58	510	540	599
6	-0.85	0.07	1.59	510	540	596
7	-0.75	0.13	1.72	510	540	603
8	-0.78	0.09	1.47	510	540	594
10	-0.58	0.27	2.00	510	540	600

Table 4-24. Science Cut Scores

Grade	Theta Cuts			Scale-Score Cuts		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
5	-0.81	0.07	1.31	510	540	582
8	-0.54	0.23	0.89	510	540	569

Table 4-25. English Language Arts, Mathematics, and Science Scaling Constants

Grade	ELA		Mathematics		Science	
	A	B	A	B	A	B
3	43.1	541.8	35.0	534.8	—	—
4	39.5	535.8	34.9	536.0	—	—
5	35.4	529.2	38.6	538.0	33.7	537.5
6	34.6	532.5	37.0	537.5	—	—
7	31.3	532.6	39.9	534.8	—	—
8	29.4	525.6	39.3	536.4	44.6	529.8
10	31.4	532.4	35.0	530.4	—	—

Note. ELA = English language arts; A = slope; B = intercept.

Properties of Scale Scores

The derived scale scores are decimal numbers and must be rounded up to the nearest integers. The IRT model cannot estimate the thetas of extreme scores (e.g., 0 and perfect raw scores) because responses to all items are identical. Software typically assigns those raw score points a theta of -99 or 99. To keep the scale score meaningful, the LOSS and the HOSS are set to cap scale scores within a reasonable range. KAP’s LOSS and HOSS are set as 400 and 700, respectively.

4.4.3. Operational Test Results

This section presents results from the 2025 KAP administration, including descriptive statistics on the number of students tested across subgroups; 2025 scale score summaries for all students and by subgroup; 2025 performance-level distributions by subject and grade; and 2025 participation data, scale-score summaries, and proficiency rates compared with previous years. Participation rates are emphasized because accounting for variability in participation is essential when interpreting KAP performance both within a given year and across years.

Student Participation

In 2025, states administered the KAP operational test in ELA and mathematics at grades 3–8 and 10, and science at grades 5 and 8. As described in Section 1.3. Required Assessments and Intended Population, Kansas is committed to including all students in the KAP assessment.

Table 4-26 shows the number of students enrolled and tested, as well as the participation rate by subject and grade. The definitions for the indicators are:

- *Enrolled students* are students assigned to take a KAP test.
- *Tested students* are students receiving a score report. Students receive a score report when they are not exempt (exemption rules are described in Section 1.3. Required Assessments and Intended Population), complete at least five items in each of the two test sections, and have logged out of the testing platform for the

first section. This reporting rule has been used since 2015.

- The *participation rate* is calculated as the number of tested students divided by the number of enrolled students.

As shown in Table 4-26, across all grades and subjects, participation in the 2025 KAP assessments ranges from 99% to 100%. Enrollment and tested counts were closely aligned, differing by fewer than 100 students per grade in most cases. English language arts and mathematics were administered in grades 3–8 and grade 10, with enrollments ranging from approximately 34,000 to 36,700 students. Science was administered in grades 5 and 8, with enrollments near 35,600 and almost full participation. The only slight decrease in participation occurred in grade 10, where rates for both English language arts and mathematics were 99%.

Table 4-26. Number and Participation Rate (PR) of Enrolled and Tested Students by Subject and Grade

Grade	English Language Arts			Mathematics			Science		
	Enrolled (N)	Tested (N)	PR (%)	Enrolled (N)	Tested (N)	PR (%)	Enrolled (N)	Tested (N)	PR (%)
3	35,205	35,161	100	35,366	35,337	100	—	—	—
4	34,433	34,399	100	34,586	34,557	100	—	—	—
5	35,452	35,406	100	35,564	35,523	100	35,599	35,556	100
6	35,417	35,346	100	35,595	35,517	100	—	—	—
7	35,753	35,668	100	35,877	35,754	100	—	—	—
8	35,610	35,518	100	35,720	35,588	100	35,767	35,683	100
10	36,707	36,492	99	36,748	36,505	99	—	—	—

Table 4-27 shows participation rates by student group and by School Board of Education (SBOE) district. The participation rates by student group and by SBOE district are not subject specific. Students are included in the calculations if they took at least one subject of the 2025 KAP assessment. The 286 school districts in Kansas are distributed among 10 SBOE districts. Some school districts appear in multiple SBOE districts when district boundaries reach into more than one SBOE district. The [Kansas Unified School Districts](#) document lists the school districts included in each SBOE district. Comparing the participation rates of students within each subject and grade by gender, ethnicity, race, EL status, and disability status, we note the following results:

- Female and male students maintained nearly identical participation patterns. Female participation ranged from 97%–99% in grades 3–8 and 96% in grade 10, while male participation followed the same trend with only minor (1%) differences by grade.
- Students in the White, Asian, and Other categories consistently participated at rates of 97%–99% across all grades.
- American Indian and Black students showed slightly lower participation in some

grades, especially by grade 10 (90–96%).

- Native Hawaiian/Pacific Islander (NHPI) participation mirrored this pattern, remaining high in early grades (96–99%) but dipping to 89% in grade 10.
- Non-Hispanic students had a slightly higher participation rate than Hispanic students in high schools.
- Non-ELs had a slightly higher participation rate than ELs in elementary schools.
- Students without disabilities had a slightly higher participation rate than students with disabilities, especially in high schools.

The comparison of participation rates of different SBOE districts within each grade showed the following results:

- Most districts maintained 99% participation through the early and middle grades, with only slight decreases in the upper grades.
- SBOE districts 1, 2, 5, and 9 showed the most consistent participation across all grades, remaining between 97% and 99% throughout.
- The lowest participation rates occurred in SBOE District 8 (declining to 92% in grade 10) and in districts 7, 9, and 10, which also showed minor declines in the upper grades (94–95%).

Districts 1 and 4 include the Kansas City, Topeka, and Lawrence school districts. District 8 includes the Wichita school district.

Table 4-27. Participation Rate by Demographic Characteristics and State Board of Education (SBOE) District

Characteristic	Grade						
	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	10 (%)
Gender							
Female	99	99	99	98	98	97	96
Male	98	99	98	98	98	97	96
Race							
Native American	98	98	97	99	97	96	96
Asian	97	97	98	98	98	98	97
Black	98	98	98	97	96	95	90
NHPI	98	98	99	97	96	98	89
Other	99	99	99	98	97	97	95
White	99	99	99	98	98	97	96
Hispanic							
Yes	99	98	98	98	98	97	95
No	99	99	99	98	98	97	96
SWD							
Yes	97	97	97	97	96	95	93
No	99	99	99	98	98	98	96
English learner							
Yes	98	97	97	96	96	95	93
No	99	99	99	98	98	98	96
Board district							
1	99	99	99	98	98	96	96
2	99	99	99	98	98	97	97
3	98	98	98	98	98	97	96
4	99	99	98	98	98	97	96
5	99	99	99	99	99	98	98
6	99	99	98	98	97	97	96
7	99	99	99	97	97	97	95
8	98	98	98	97	96	96	92
9	99	99	99	98	97	97	94
10	99	99	99	97	97	97	94

Note. NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability.

For all tested students, Table 4-28 shows the percentage of students in each student group by grade. This summary is not subject specific. If a student tested in one subject of the KAP assessment, then the student is included in the calculation. The student groups include gender, race, ethnicity, disability status, and EL status.⁵ The percentages of students in

⁵ Economically disadvantaged status is not shared with AAI to protect the privacy of students, so this student group is not included in the comparison.

each student group were very similar across grades, except students with disabilities. There were approximately equal percentages of male and female students. The largest percentage tested by race group was White, and the largest percentage tested by ethnicity group was non-Hispanic. More students without disabilities were tested than students with disabilities, and more non-ELs were tested than ELs. There was a decrease in the percentage of students with disabilities across grades. The lower grades had higher percentages of students with disabilities than did higher grades.

Table 4-28. Percentage of Tested Students by Demographic Characteristic and Grade

Characteristic	Grade						
	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	10 (%)
Gender							
Female	49	49	49	49	49	49	49
Male	51	51	51	51	51	51	51
Race							
Native American	2	2	2	2	2	2	2
Asian	3	3	3	3	3	3	3
Black	7	7	7	7	7	7	8
NHPI	0	0	0	0	0	0	0
Other	8	8	8	8	7	7	7
White	80	80	80	80	80	80	80
Hispanic							
Yes	23	23	23	22	22	22	23
No	77	77	77	78	78	78	77
SWD							
Yes	17	17	17	16	15	14	12
No	83	83	83	84	85	86	88
English learner							
Yes	15	14	13	12	11	10	9
No	85	86	87	88	89	90	91

Note. NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability.

Operational Test Results

Table 4-29, Table 4-30, and Table 4-31 present summaries of scale scores by grade for ELA, mathematics, and science. As noted previously, it is critical to consider variability in participation rates when interpreting KAP performance within and across years.

The minimum and maximum scale scores for each grade and subject were 400 and 700, respectively. As shown in Table 4-29, Table 4-30, and Table 4-31, scale-score distributions for ELA, mathematics, and science were generally stable and consistent across grades, with only modest declines in mean performance from lower to upper grades. ELA means decreased from 545 in grade 3 to 530 in grade 10, mathematics from 535 in grade 3 to 528 in grade 10, and science from 536 in grade 5 to 525 in grade 8. Standard deviations declined with grade level in all subjects, indicating reduced score variability among older

students. Median and percentile values showed parallel trends, with slightly narrower score ranges at higher grades, reflecting greater score concentration. All subjects shared the same fixed scale of 400–700, with no evidence of ceiling or floor effects, suggesting appropriate use of the reporting scale and consistent measurement across grades and subjects.

Table 4-29. Scale-Score Descriptive Statistics for English Language Arts

Grade	M	SD	Min	P₁₀	P₂₅	P₅₀	P₇₅	P₉₀	Max
3	545.3	51.6	400	484	505	540	577	609	700
4	541.6	50.3	400	480	505	540	572	607	700
5	539.2	44.3	400	481	506	540	572	600	700
6	531.8	42.6	400	477	502	531	560	585	700
7	532.7	42.2	400	480	503	530	558	590	700
8	532.4	41.0	400	480	500	528	558	587	700
10	530.3	39.7	400	481	500	528	554	580	700

Note. M = mean; SD = standard deviation; P₁₀, P₂₅, P₅₀, P₇₅, and P₉₀ = 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

Table 4-30. Scale-Score Descriptive Statistics for Mathematics

Grade	M	SD	Min	P₁₀	P₂₅	P₅₀	P₇₅	P₉₀	Max
3	535.2	40.4	400	488	506	530	560	589	700
4	536.9	41.6	409	490	507	531	558	587	700
5	538.2	44.0	400	491	506	533	560	599	700
6	535.3	44.2	412	489	502	529	557	587	700
7	532.4	46.6	400	482	499	525	556	595	700
8	534.6	46.4	406	486	501	527	559	594	700
10	528.4	39.6	400	484	505	524	551	580	700

Note. M = mean; SD = standard deviation; P₁₀, P₂₅, P₅₀, P₇₅, and P₉₀ = 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

Table 4-31. Scale-Score Descriptive Statistics for Science

Grade	M	SD	Min	P₁₀	P₂₅	P₅₀	P₇₅	P₉₀	Max
5	536.4	37.1	400	489	510	536	560	586	700
8	524.8	50.6	400	463	486	525	559	592	700

Note. M = mean; SD = standard deviation; P₁₀, P₂₅, P₅₀, P₇₅, and P₉₀ = 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

Table 4-32, Figure 4-4, Figure 4-5, and Figure 4-6 provide the percentage of students achieving each performance level (i.e., level 1 through level 4) and the proficiency rate (i.e., percentage at level 3 and level 4) by subject and grade.

In ELA, the percentage of students performing at level 1 ranged from 26% (grade 3) to a high of 29% (grades 8 and 10), while level 4 performance was highest in grade 3 (13%) and lowest in grades 6–7 (8–9%). Mid-level performance remained dominant, with level 3 peaking at 40% in grade 3 and lowest at 29% in grade 10.

For mathematics, level 1 percentages increased with grade, from 23% in grade 3 to a high of 32% in grade 10, while level 4 declined steadily from 20% in grade 3 to just 5% in grade 10—the lowest high-performance rate among all subjects. Level 3 peaked at 33% in grade 5 and was lowest at 23% in grade 3, showing most students concentrated in levels 2–3 across grades.

In science (grades 5 and 8 only), level 1 rose from 23% in grade 5 to 34% in grade 8, while level 4 dropped from 12% to 20%—notably the only subject where the higher grade had a *higher* proportion at level 4. Level 3 was highest at 34% in grade 5 and lowest at 19% in grade 8, with total proficiency (levels 3–4) decreasing from 46% to 39% between grades.

Table 4-32. Percentage of Students Achieving at Each Performance Level (PL) for ELA, Mathematics, and Science

Grade	ELA PL (%)					Mathematics PL (%)					Science PL (%)				
	1	2	3	4	P	1	2	3	4	P	1	2	3	4	P
3	26	21	40	13	53	28	29	23	20	43	—	—	—	—	—
4	27	23	39	12	51	27	30	28	15	43	—	—	—	—	—
5	27	23	40	11	51	27	30	33	10	43	23	31	34	12	46
6	27	30	33	9	42	26	34	31	10	41	—	—	—	—	—
7	28	29	36	8	44	27	36	29	8	37	—	—	—	—	—
8	29	30	33	9	42	27	34	28	11	39	34	27	19	20	39
10	29	32	29	10	39	32	34	29	5	34	—	—	—	—	—

Note. P = proficiency (combination of performance levels 3 and 4). Column percentages may not total 100 because of rounding.

Figure 4-4. Performance-Level Distribution for English Language Arts

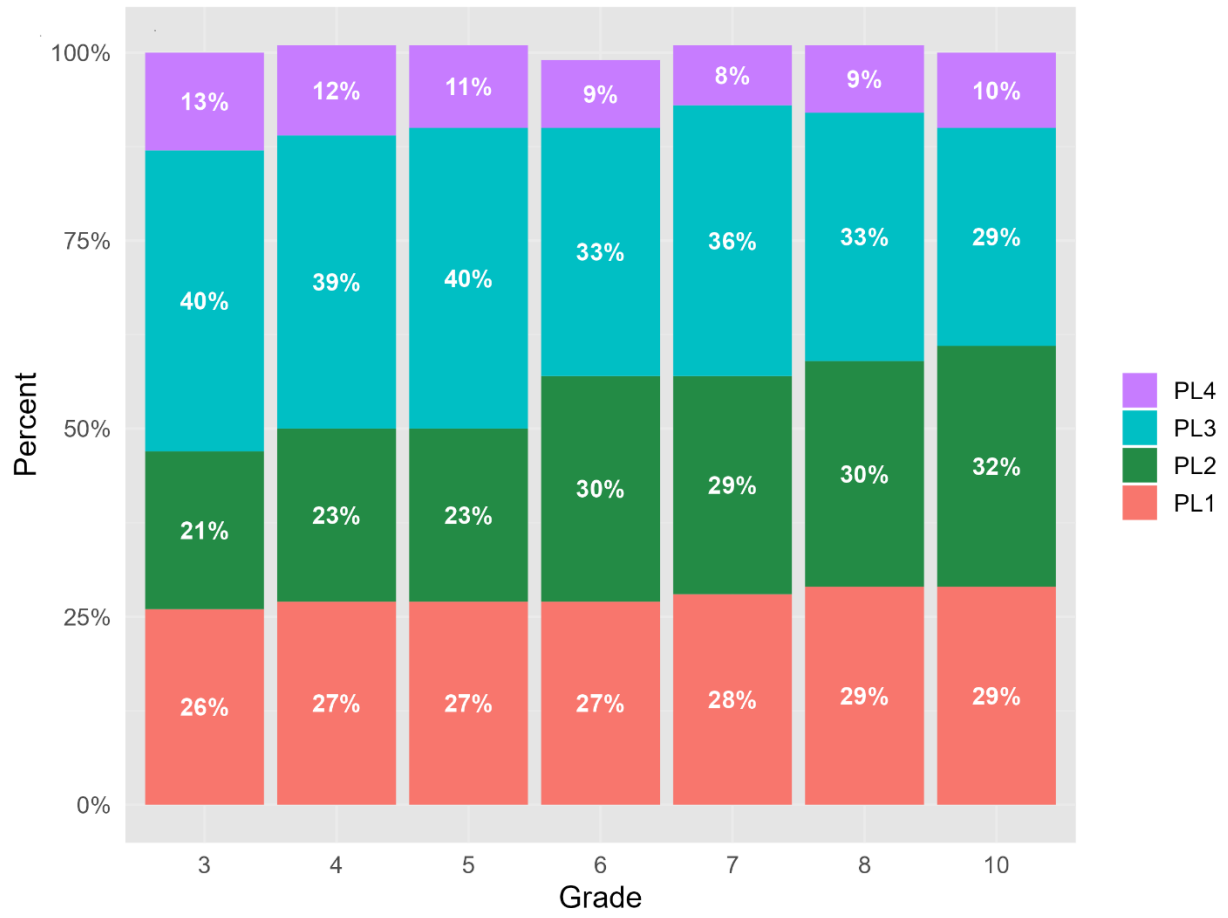


Figure 4-5. Performance-Level Distribution for Mathematics

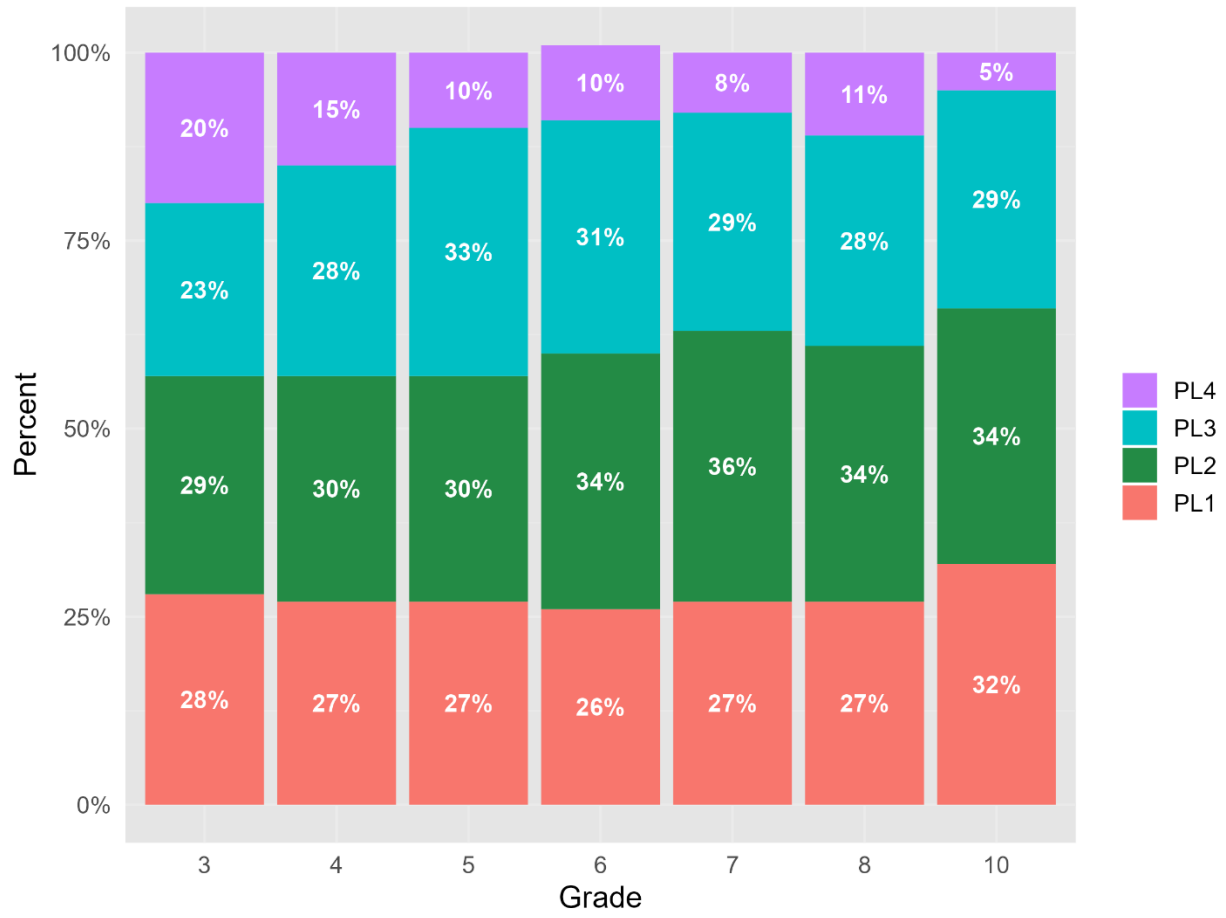


Figure 4-6. Performance-Level Distribution for Science

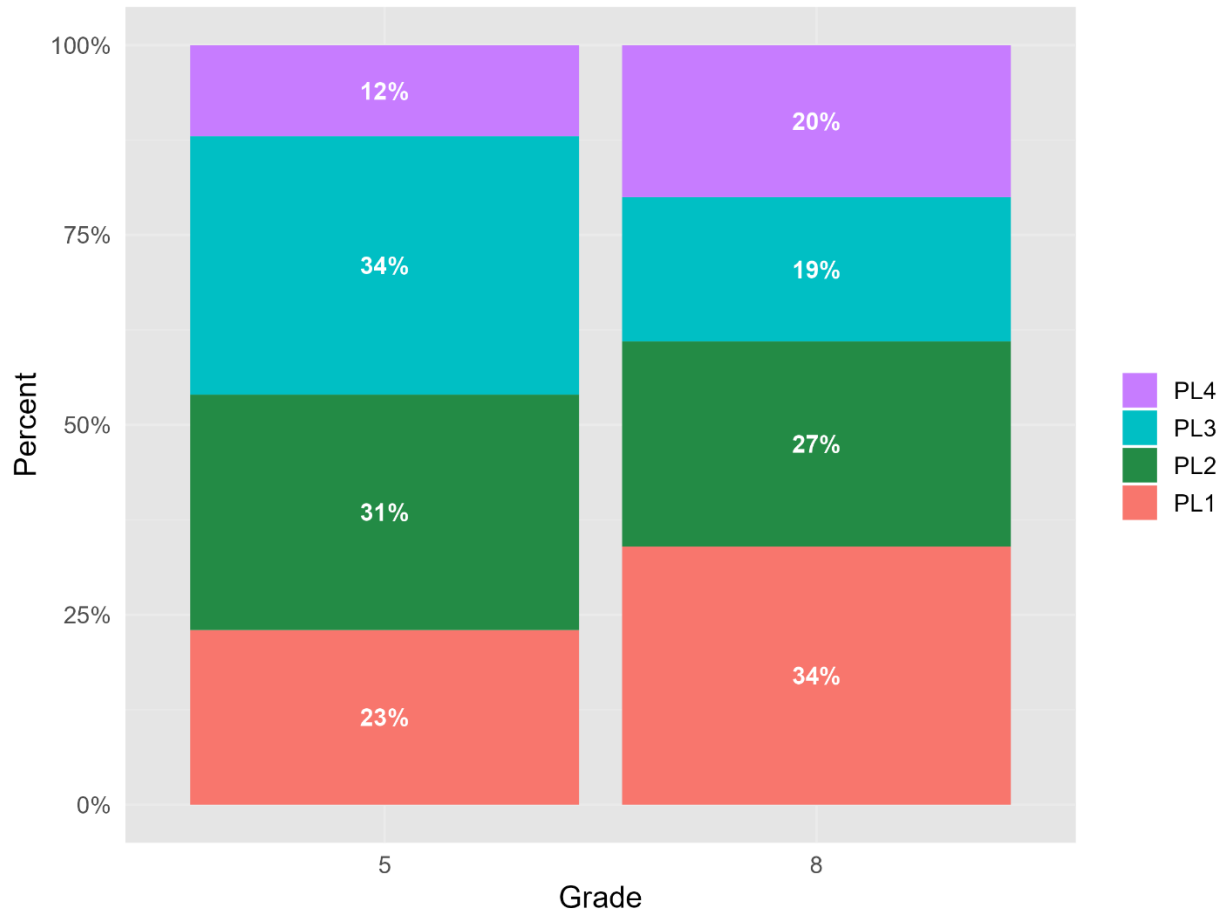


Table 4-33, Table 4-34, and Table 4-35 summarize the mean and standard deviation of the scale scores by demographic student group.⁶ For the 2025 KAP assessments in ELA, mathematics, and science, patterns were consistent across content areas, with the largest mean differences observed by EL and student-with-disability status, moderate differences by race/ethnicity, and smaller differences by gender. Across all subjects, performance gaps tended to narrow slightly at higher grades, and standard deviations decreased modestly, indicating reduced score variability among older students.

⁶ Economically disadvantaged status is not shared with ATS to protect the privacy of students, so this student group is not included in the comparison.

Table 4-33. English Language Arts Mean and Standard Deviation of Scale Scores by Grade and Student Subgroup

Characteristic	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Gender														
Male	541.8	51.4	539.8	50.9	536.7	44.6	528.0	42.2	530.2	42.9	527.8	40.8	526.6	39.8
Female	549.0	51.5	543.5	49.6	541.8	43.8	535.7	42.7	535.4	41.3	537.2	40.7	534.2	39.1
Race														
NA	523.9	43.7	521.8	44.7	521.1	39.7	512.5	38.9	514.8	36.6	516.5	35.1	515.3	32.6
Asian	556.9	52.0	555.5	54.6	552.0	47.0	543.0	44.2	546.8	47.4	550.6	45.8	549.2	45.8
Black	522.9	44.6	516.2	44.6	518.7	41.2	509.8	38.5	510.4	38.2	513.1	34.2	513.8	34.4
NHPI	526.2	45.6	538.1	48.1	518.0	43.2	515.5	36.7	520.3	37.0	515.4	34.2	512.7	37.7
Other	541.7	51.0	536.5	49.4	533.6	43.3	526.2	41.3	529.7	41.2	529.7	42.1	527.3	38.4
White	547.7	51.7	544.3	50.0	541.6	43.9	534.3	42.3	534.9	41.8	534.2	40.8	531.8	39.5
Hispanic														
Yes	526.0	45.3	523.9	44.9	521.6	40.4	514.9	38.3	516.0	38.1	517.9	35.5	516.3	35.1
No	550.9	51.9	546.7	50.6	544.2	44.0	536.6	42.5	537.4	42.1	536.5	41.6	534.4	40.0
SWD														
Yes	512.7	46.3	506.9	46.3	504.9	39.0	499.0	37.8	499.9	36.1	500.9	33.1	499.9	32.4
No	552.1	50.0	548.7	48.1	546.0	42.0	537.9	40.6	538.3	40.6	537.5	39.9	534.5	38.7
EL														
Yes	516.2	41.1	512.9	40.6	508.9	35.7	499.4	32.2	498.0	31.3	499.5	27.2	495.7	26.8
No	550.2	51.6	546.1	50.2	543.5	43.7	535.9	42.0	536.7	41.5	535.7	40.7	533.5	39.1

Note. M = mean; SD = standard deviation; NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

Table 4-34. Mathematics Mean and Standard Deviation of Scale Scores by Grade and Student Subgroup

Characteristic	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Gender														
Male	539.1	42.3	541.7	44.5	541.4	46.8	538.3	46.3	535.5	49.5	536.4	48.4	530.6	42.0
Female	531.1	37.9	532.0	37.8	534.8	40.7	532.2	41.6	529.1	43.0	532.6	44.2	526.2	36.9
Race														
NA	520.8	33.9	520.7	33.4	521.5	33.2	517.3	34.7	513.4	36.4	516.1	36.9	511.6	30.7
Asian	549.9	47.4	554.5	53.4	559.8	54.4	556.9	56.3	560.2	61.6	564.9	60.8	559.0	56.6
Black	512.9	33.6	514.1	33.1	515.7	33.6	510.6	31.7	507.2	34.3	509.9	33.0	508.0	30.6
NHPI	518.5	36.9	528.7	38.9	521.4	37.8	518.8	37.8	519.8	47.4	519.0	37.9	511.4	32.2
Other	529.5	39.3	530.9	39.6	530.9	42.2	527.6	41.7	525.6	41.5	529.2	46.0	522.4	37.5
White	537.5	40.2	539.3	41.3	540.6	43.9	537.9	44.0	534.7	46.4	536.6	45.9	530.1	38.9
Hispanic														
Yes	519.0	33.7	520.7	34.0	521.5	35.2	518.1	34.7	514.1	36.2	517.0	37.8	512.3	31.9
No	539.9	41.0	541.6	42.4	543.0	45.2	540.2	45.3	537.6	47.9	539.6	47.4	533.2	40.4
SWD														
Yes	512.8	36.9	513.4	37.0	511.6	36.9	506.6	35.9	502.7	37.2	503.5	34.7	501.6	31.7
No	539.8	39.6	541.7	40.9	543.5	43.4	540.7	43.5	537.4	46.1	539.6	46.1	532.1	39.2
EL														
Yes	514.4	32.5	515.0	32.0	515.0	32.0	507.3	29.8	502.2	30.1	501.8	28.8	500.1	27.4
No	538.8	40.6	540.5	41.9	541.6	44.5	539.0	44.4	536.0	46.9	538.0	46.6	531.1	39.6

Note. M = mean; SD = standard deviation; NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

Table 4-35. Science Mean and Standard Deviation of Scale Scores by Grade and Student Group

Characteristic	Grade 5		Grade 8	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Gender				
Male	539.2	38.8	526.4	52.8
Female	533.4	34.9	523.2	48.1
Race				
NA	522.1	31.6	503.4	46.2
Asian	544.7	39.6	543.1	56.2
Black	516.1	32.2	495.0	43.0
NHPI	515.0	34.3	502.2	42.0
Other	531.3	35.8	520.7	50.2
White	538.8	36.9	527.8	50.0
Hispanic				
Yes	521.2	33.0	502.9	46.6
No	540.8	37.0	531.1	49.9
SWD				
Yes	514.2	36.2	490.2	45.5
No	540.8	35.6	530.4	49.1
EL				
Yes	512.3	30.9	479.3	39.0
No	540.0	36.6	529.6	49.2

Note. *M* = mean; *SD* = standard deviation; NA = Native American; NHPI = Native Hawaiian and Pacific Islander; SWD = student with disability; EL = English learner.

Participation Trend

Table 4-36 presents enrollment trends for 2021–2025 for ELA, mathematics, and science.

Table 4-36. Total Number of Enrolled Students by Subject and Grade for 2021–2025

Subject	Grade	2021	2022	2023	2024	2025
English language arts	3	35,440	35,356	35,503	34,583	35,205
	4	35,547	35,878	35,595	35,522	34,433
	5	36,735	35,799	36,053	35,701	35,452
	6	37,225	36,953	35,899	36,118	35,417
	7	38,145	37,370	37,096	36,002	35,753
	8	38,275	38,173	37,579	37,290	35,610
	10	36,811	36,747	38,184	38,072	36,707
Mathematics	3	35,455	35,389	35,562	34,739	35,366
	4	35,557	35,907	35,648	35,643	34,586
	5	36,743	35,830	36,095	35,799	35,564
	6	37,224	36,968	35,946	36,213	35,595
	7	38,142	37,387	37,147	36,087	35,877
	8	38,286	38,191	37,645	37,399	35,720
	10	36,813	36,799	38,225	38,152	36,748
Science	5	36,756	35,849	36,104	35,815	35,599
	8	38,301	38,204	37,650	37,410	35,767

Figure 4-7 presents the participation rates (i.e., proportion of students receiving a score report out of students enrolled) for different subjects and grades by year from 2021–2025. Across all subjects, total student enrollment in KAP remained stable from 2021 through 2025, with only modest year-to-year variation. In ELA, grade-level enrollments ranged from approximately 34,000 to 38,000 students annually, showing a gradual decline over time—most notably in grades 7 and 8, which decreased by roughly 2,500 students since 2021.

Mathematics followed nearly identical enrollment patterns, with parallel grade-level counts across years, reflecting consistent participation rates across content areas.

Science, administered in grades 5 and 8, also showed slight decreases in enrollment from 2021 to 2025, aligning with overall statewide enrollment trends. Across all grades and subjects, total enrollment declined by roughly 5–7% from 2021 to 2025, indicating a small but steady reduction in the assessed population over the five-year period.

Figure 4-7. Participation Rates for 2021–2025 by Subject and Grade

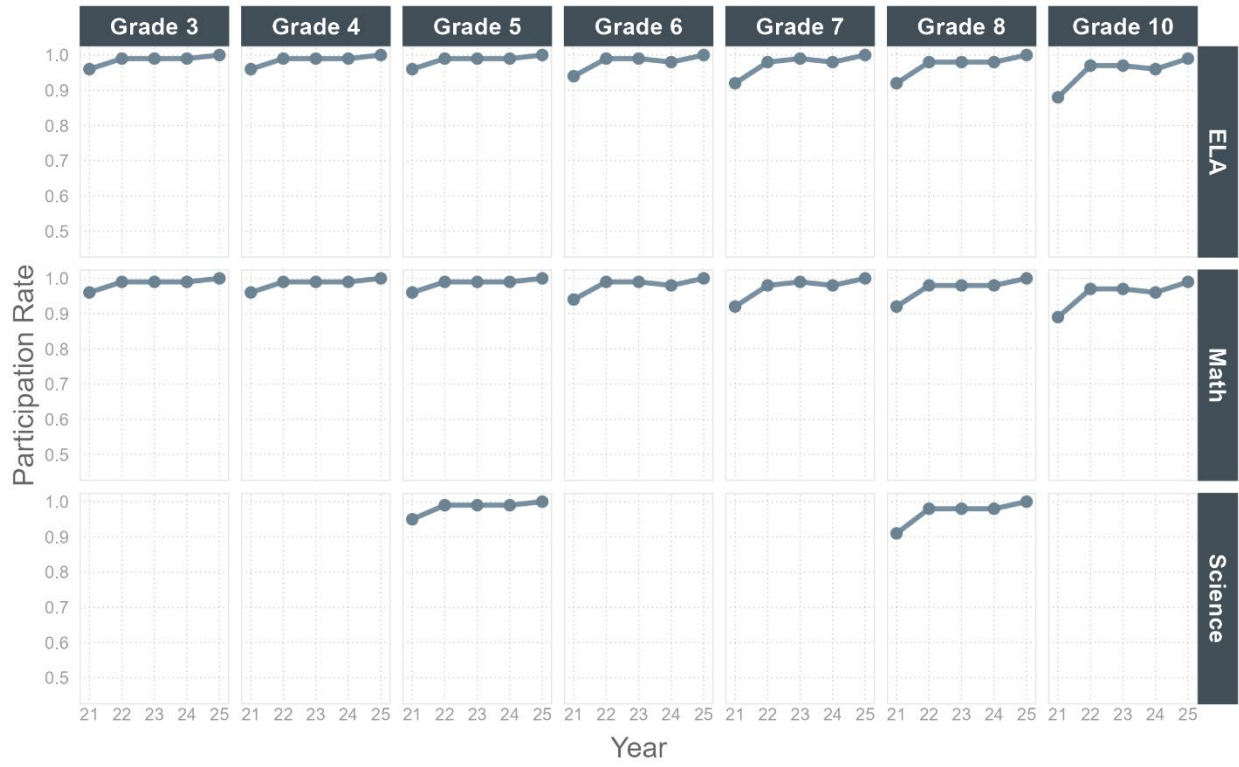


Table 4-37. Proficiency Rates for English Language Arts, 2021–2025

Year	Grade						
	3	4	5	6	7	8	10
2021	40	48	44	36	32	25	28
2022	38	43	40	33	32	21	26
2023	29	44	41	35	33	22	27
2024	39	44	40	35	33	23	28
2025	53	50	50	43	43	42	39

Table 4-38. Proficiency Rates for Mathematics, 2021–2025

Year	Grade						
	3	4	5	6	7	8	10
2021	48	33	29	28	24	22	21
2022	50	35	31	30	27	21	21
2023	51	39	33	32	29	24	22
2024	51	38	33	31	29	25	22
2025	43	43	43	41	37	39	35

Table 4-39. Proficiency Rates for Science, 2021–2025

Year	Grade	
	5	8
2021	44	28
2022	44	23
2023	44	24
2024	46	26
2025	46	39

Quality-Control Checks

The scoring and reporting process of KAP test results had multiple quality-control steps. First, student response data were checked at least three times during the testing window for scoring errors or duplicates.

Second, we conducted classical item analysis during the testing window using approximately 20% of the overall test volume. We compared classical item statistics from the 2025 operational administration with classical item statistics from previous administrations. The purpose of this step was to monitor classical item statistics trends and ensure items were functioning as expected. Third, two psychometric staff members independently generated and compared scoring tables. They examined reasonableness and accuracy of the scoring tables through predetermined criteria:

- All subjects and grades were represented.
- All tests were represented.
- All raw scores were represented for each form.
- No integer was missing from the scale scores, from 0 to the maximum test-form score.
- The scale score increased with the raw score within each form.
- The minimum scale score was 400, and the maximum scale score was 700.

Fourth, at least two psychometric staff members independently checked the cut scores used to classify students to ensure they were consistent with the cut scores approved by the Kansas SBOE.

Finally, the psychometric and technology teams independently calculated each individual student's total score, scale score, performance levels, subscores, and subscore performance levels. We compared results from the two teams' independent calculations to identify any differences or calculation errors. We generated students' score reports only after the scoring results from both teams were identical. The purpose of all quality-control steps was to ensure the scoring results provided on students' reports were complete and accurate.

4.5. Multiple Assessment Forms

In large-scale assessment programs, different item sets may be used on test forms within and across years. Linking the scores from these different test forms puts the form scores on a common scale and ensures that all forms for a given grade and subject area provide comparable scores. This outcome means that students will not have an unfair advantage or disadvantage simply because they took an easier or harder test form than other students did.

The 2025 administration is the first operational administration for the updated KAP. Only one operational form was developed and administered; thus, no linking was conducted.

4.5.1. Cross-Year Linking Design

To increase the number of linking items and maximize linking stability, the cross-year linking uses the preequating method. When the items from different years are on the same IRT scale, the student scale scores calculated from these IRT item parameters are equated and placed on the base scale.

4.5.2. Cross-Year Linking Procedure

All items on the 2025 ELA, mathematics, and science tests were field tested in previous years. In those years, all field-tested items were calibrated using concurrent item calibration after the test window closed, by fixing the item parameters of the operational items so that the field-test item parameters were placed on the same IRT scale as operational items, that is, the base scale. Also, the test characteristics such as TIF and

scoring tables are compared across years during test-construction psychometric review to ensure that different test forms across years have similar test characteristics, that is, similar reliability estimates and similar raw-score cuts.

4.6. Multiple Versions of an Assessment

The KAP assessment is administered online via the Kite platform, which can be used on PCs with Windows, Macs, Chromebooks, and iPads. All students who take the KAP assessment must use the Kite Student Portal (described in Section 2.3.2. Test Administration Procedures).

The Kite platform can provide various accommodations for students with special needs. For details about available accommodations, please refer to Section 5.4. Accommodations. The one exception is that a paper-pencil braille form is provided to students who need it. No grade or subject-area test has more than 10 students taking the braille form.⁷ The braille version has the same operational items as the online version but no field-test items. When Allyant translated items to braille format, it modified some formats of items to provide an adequate experience for students who are blind or visually impaired without introducing construct-irrelevant variance. For example, the radio buttons of the selected-response items on the online version are changed to option labels (i.e., A, B, C, and D). Moreover, Allyant and the AAI content team collaborate to construct the test administration notes for the braille form, which add clarifying language so that students who are blind or visually impaired can access the same information as their sighted peers.

4.7. Technical Analysis and Ongoing Maintenance

This technical manual includes a series of technical analyses that use this year's testing data. These analyses include classical item analyses, DIF analysis, IRT calibration and scaling, relationships among different assessments, reliability analyses, classification consistency and accuracy analyses, and test result summaries.

Since this is the first test administration for the updated KAP assessments, we plan to monitor relationship between KAP results and NAEP results to establish the validity of the KAP assessment. An independent alignment study is also planned for the upcoming school year to evaluate the alignment between KAP assessments and the Kansas Standards.

⁷ The sample sizes of braille forms were too small to undertake a comparability study between the braille version and online version.

Chapter 5 | Inclusion of All Students

This chapter presents information about KAP’s inclusion of all students, including students with disabilities and English learners (EL). The procedures for including students with disabilities and ELs are summarized, followed by a description of the available accessibility tools and accommodations. More information about accessibility supports and accommodations for KAP can be found in the [Kansas Accessibility Manual](#), the [Kite Accessibility Manual](#), and the [KAP Test Coordinator Manual](#).

The Kansas State Department of Education (KSDE) complies with the Elementary and Secondary Education Act (ESEA) and the Individuals with Disabilities Education Improvement Act (IDEA), both of which require all students, including students with disabilities and ELs, to participate in assessments used for accountability purposes. One of the principles of ESEA is strong accountability for educational achievement results for all students. Through this federal legislation, assessments that aim to increase accountability provide important information regarding (a) schools’ success in including all students in standards-based education, (b) students’ achievement of standards, and (c) improvements needed for specific groups of students. IDEA explicitly governs services provided to students with disabilities. Accountability at the individual level is provided through the Individualized Education Program (IEP), Section 504 plan, or individual learning plan (ILP). All of those plans are developed to address each student’s unique needs.

5.1. Procedures for Including Students with Disabilities

Accessibility tools and accommodations that are available either within or outside the Kite® system allow students with disabilities to take KAP assessments. The details about different tools and accommodations are in Section 5.3. Accessibility Tools and Section 5.4. Accommodations. The inclusion of students with disabilities is achieved by providing clear guidelines for educators, so they can register their students with different needs. The [KAP Test Coordinator Manual](#) describes step-by-step registration procedures for students who need accommodations.

5.2. Procedures for Including English Learners

As described in Section 1.3. Required Assessments and Intended Population, ELs are required to take KAP assessments with the exception of English language arts (ELA) in the first year. Accessibility tools and accommodations that are available either within or outside the Kite system allow ELs to take KAP assessments. Specific accessibility tools and accommodations for ELs include directions read aloud by a synthetic voice, electronic translators and word-to-word translators (not for ELA passages), translation dictionaries, and Spanish key word translation for mathematics and science assessments. Details about different tools and accommodations are in Section 5.3. Accessibility Tools and Section 5.4. Accommodations. The inclusion of ELs is achieved by providing clear guidelines for educators, so they can register their students with different needs. The [KAP Test Coordinator Manual](#) describes step-by-step registration procedures for students who

need accommodations.

5.3. Accessibility Tools

Accessibility tools are available for all students taking KAP assessments. Accessibility tools available for students vary by subjects. Table 5-1 describes the accessibility supports available to all students who take KAP assessments, as well as their descriptions and recommendations for use.

Table 5-1. Accessibility Tools for KAP

Tool	Description
Calculator: Basic four-function	Allows students to perform the basic four functions: addition, subtraction, multiplication and division. This tool is available for mathematics grades 6–8 and 10, and science grade 5 in calculator-allowed sections.
Calculator: TI graphing	Allows students to plot graphs, solve equations, and display several lines of calculations on the screen. This tool is available for mathematics grade 10. May not be available in mathematics sections that measure numbers and operations.
Calculator: TI scientific	Allows students to perform calculations in science, engineering, and mathematics. This tool is available for mathematics grades 6–8 and science grades 8 and 11. May not be available in mathematics sections that measure numbers and operations.
Math resource sheet	Presents a PDF document to students that includes grade-appropriate mathematics formulas. There are resource sheets available for mathematics grades 4–8 and 10.
Eraser	Removes highlighting and striker marks from the screen.
Guide line	When selected, follows the student’s pointer and highlights the text of a reading passage line by line. This tool differs for iPads, where the line remains stationary as the student scrolls through the passages.
Highlighter	Allows students to select text on the screen and highlight the selected text with a pink background.
Mark for review: question answered	When selected by test takers, changes the item number indicator at the top of the screen to blue with an accompanying flag graphic.
Mark for review: question unanswered	When selected by test takers, changes the item number indicator at the top of the screen to red with an accompanying flag graphic.
Notes	Presents a yellow rectangle on the screen where students can type notes about the test content. This tool is only available for ELA tests.

Tool	Description
Periodic table	Presents a standard periodic table. Students can select an individual element to view the atomic number, atomic mass, and full element name. The default view shows elements by their abbreviations. This tool is available for the grade 11 science test.
Pointer	Allows students to select items in the test.
Search	Allows students to enter search terms; matching words are then highlighted in orange. This tool is available for ELA and science tests.
Striker	Allows students to place a line through an answer choice that is not desired.
Tags	Allows students to use various tags within a reading passage. The available tags are Main Idea, Supporting Detail, Key Word, Evidence, Reread This, and “Is this Important?”
TTS: directions	Allows students to choose to have a synthetic voice read directions aloud on the assessment.
TTS: mathematics and science	Allows students to choose to have a synthetic voice read directions, stimuli, and test items aloud on the mathematics and science assessments.
Whole-screen magnification	Allows students to magnify the screen up to four levels.
Sketch pad	Allows students to draw, write, create shapes, etc.

Note. ELA = English language arts; TTS = text-to-speech audio.

5.4. Accommodations

Assessment accommodations are practices and procedures that provide equitable access during instruction and assessments for students with special needs. These accommodations may not alter the assessment’s validity, score interpretation, reliability, or security. They are designed to reduce or eliminate the effects of a student’s disability or English proficiency; however, they do not alter learning expectations. The accommodations provided to a student, documented in a student’s IEP, Section 504 plan, or ILP, should be the same for classroom instruction, classroom assessments, and local education agency and state assessments.

It is critical to note that some accommodations that are appropriate for instructional uses may not be appropriate for use on standardized assessments. For example, a student with low vision will need accommodations to make a test accessible. However, in an ELA assessment, reading passages aloud to a student would change what is being measured and is therefore not a valid accommodation. Use of a magnifying tool or a large-print version of a test is an acceptable accommodation.

It is important for educators to become familiar with state policies regarding

accommodations during assessments. According to the [KAP Summative Test Administrator Manual](#) (p. 24), reading any text (including isolated words) to students in the passages on the ELA test is prohibited. Only a very limited number of students, such as those who cannot access printed text, may be permitted to have passages read through text-to-speech (TTS) software with approval from KSDE staff. Another prohibited accommodation is for teachers and students to bring pre-generated journals and logs.

The [KAP Summative Test Administrator Manual](#) provides more details regarding accommodations in KAP assessments, including an overview, prohibited practices, and recording accommodations used during testing (e.g., most testing accommodations should be entered into the student's Personal Needs and Preferences [PNP] Profile). Additional information about accommodations or Kite tools can be found in the [Kansas Accessibility Manual](#) or [Kite Accessibility Manual](#). Table 5-2 presents the accommodations available for KAP assessments.

Table 5-2. Accommodations Available for KAP Assessments

Tool	Description
American Sign Language (ASL)	Displays available ASL videos for the assessment question. This accommodation is only available for mathematics and science.
Auditory calming	Provides relaxing, peaceful music that can play while the student takes the test.
Braille form ⁸	Provides a paper-pencil braille test form.
Color contrast	Sets a text color and a background color. Options are gray text on black background, yellow text on black background, green text on white background, and red text on white background.
Color overlay	Provides a color background behind the content on the screen. Color options are light blue, light yellow, light gray, light red, and light green.
Key word translation	Provides Spanish translations for the key words on the items. This accommodation is only available for mathematics and science.
Masking (student-controlled or presented by default)	Allows a student to mask, or cover, parts of the test. After a student selects the masking button, a black box appears. The student can move the masking box by dragging it to different areas of the screen.
Reverse contrast	Sets the text color to white and the background color to black.
Switches	Allows students to interact with the assessments through the use of a single switch or key instead of a mouse.
Text-to-speech	Provides a synthetic voice that reads text and test items aloud on the assessment.
Whole-screen magnification	Allows students to magnify the screen according to what has been set up in their Personal Needs and Preferences (PNP) Profile.

Note: ELA = English language arts.

5.4.1. Selection of Accommodations

A few basic rules apply to every available accommodation on the KAP assessment. First and foremost, only accommodations that have been used regularly in instruction may be used on the KAP assessments. Second, students with IEPs, Section 504 plans, or ILPs may use only the accommodations documented in their plans. Finally, for accommodations to be available during the KAP assessment, teachers must submit accommodation requests

⁸ Starting in 2021–2022, a new braille online form fully aligns with the paper-pencil braille form.

through the student’s PNP Profile in Kite Educator Portal before beginning any assessment. For TTS software requests, local test administrators need to enter the support in the Accommodations section of the student’s PNP Profile. The [KAP Test Coordinator Manual](#) lists the steps for creating a PNP Profile for students.

Test administrators handle some accommodations (e.g., braille, magnification device) that are allowed for the KAP assessment, but most accommodations (e.g., color contrast) are built-in features in the Kite system. Because features in the Kite system are activated according to students’ needs, teachers are required to mark those needs in the PNP Profile. Additionally, teachers need to report in advance if braille is needed. For additional accommodations documented in a student’s plan that are not available for KAP assessments, teachers should contact the District Test Coordinator (DTC). Then, the DTC will send the request to KSDE staff for approval. These additional accommodations requested should not change the construct being tested.

5.4.2. Frequency of Accommodation Use

The summary of PNP accommodation requests shown in Table 5-3 indicates the number of students for whom each accommodation is requested. This table summarizes PNP selections by grade. Note that some students may receive multiple accommodations. The table shows that test-to-speech audio is the most commonly requested accommodation option. This accommodation makes an audio recording of the test item available.

Table 5-3. Frequency of Accommodation Requests by Grade

Accommodation	Grade						
	3	4	5	6	7	8	10
ASL	9	15	19	17	15	24	14
Auditory calming	68	151	131	87	95	271	273
Braille form	1	3	3	1	3	2	3
Color contrast	11	11	11	29	25	25	17
Color overlay	11	8	17	27	16	25	22
Key word translation	240	372	389	543	571	557	581
Masking	15	20	14	29	20	33	37
Reverse contrast	5	6	2	4	7	3	4
Switches	0	7	4	4	3	3	4
TTS	506	492	505	3,306	3,148	2,859	1,645
WSM	35	55	41	76	72	45	59
Total	901	1,140	1,136	4,123	3,975	3,847	2,659

Note: ASL = American Sign Language; TTS = text-to-speech audio; WSM = whole-screen magnification.

Chapter 6 | Academic Achievement Standards and Reporting

This chapter describes updates related to achievement standards and reporting for the Kansas Assessment Program (KAP). The achievement standards for the 2025 KAP assessments are updated to reflect the Kansas content standards and the refreshed 2025 KAP test blueprints for all three content areas. The following sections describe standard-setting procedures and outcomes for English language arts (ELA) and mathematics grades 3–8 and 10, as well as science grades 5 and 8.

6.1. State Adoption of Academic Achievement Standards for All Students

Policy performance level descriptors (PLDs) define the KAP academic achievement standards. Although the KAP assessment is based on content standards, the assessment evaluates student performance using academic achievement standards. PLDs describe the expected academic achievement at each performance level. Classifying student performance on the assessment into a given performance level means that the student meets the minimum expected knowledge and skills of that performance level. This score interpretation applies to all students who participate in the KAP assessment. The policy PLDs have four levels: level 1, level 2, level 3, and level 4. Performance levels 3 and 4 are considered meeting grade-level expectations, i.e., meeting the proficiency. The state adopted the same policy academic achievement standards across grades and subject in 2025.

6.2. Achievement Standard Setting

For the KAP assessment, achievement standard setting occurred in summer 2025 for ELA and mathematics grades 3–8 and 10 and science grades 5 and 8. The following sections describe the standard-setting method, procedure, and outcomes for all subjects and grades.

6.2.1. Standard-Setting Method

Panelists used the Bookmark standard-setting procedure (BSSP) to establish cut scores for all subjects and grades. The BSSP is widely used in K–12 educational assessment contexts. Based on panelists’ review of collections of test items, the BSSP can generate cut scores (Cizek & Bunch, 2007). In this method, according to empirical item data (e.g., item response theory [IRT] item parameter estimates), an ordered item booklet (OIB) displays items ranked from easiest to hardest. Panelists review the items in order and place a bookmark at the page in the OIB to indicate where they believe the “just-barely” examinee (i.e., minimally competent examinee or just-qualified candidate) would have a specific probability (i.e., response probability) of answering the item correctly.

Taking advantage of IRT scaling, the BSSP places students and items on the same scale.

According to the assumptions of the IRT model, a student's test score can provide a theoretically known probability for the student answering a dichotomous item (e.g., multiple-choice item) correctly. In the case of polytomously scored items, responses are assigned a given score point. The student scores can be used to rank items.

According to Cizek & Bunch (2007), the BSSP is widely used for several reasons. First, from a practical standpoint, the method can be used for complex, mixed-format assessments, and panelists using the method can consider selected-response and technology-enhanced items together. Second, for those who must make judgments regarding cut scores, the method presents a relatively simple task for participants. Third, the BSSP is also comparatively easy for those who must implement the procedure. Finally, the method has certain psychometric advantages because of its basis in IRT analysis and its fidelity to test construction techniques used during assessment development.

Given that KAP assessments are administered to a reasonably large population of students, and that BSSP was used in previous KAP programs to set performance cut scores, the BSSP was determined, in consultation with the Kansas Department of Education (KSDE) and the state's Technical Advisory Committee (TAC), to be a reasonable method for establishing cut scores.

One key element of BSSP is the OIB. The OIB contains both dichotomously scored items (e.g., multiple choice) and polytomously scored items (e.g., technology enhanced). Each dichotomously scored item appears once in the OIB in a location determined by the response probability (set as 0.67 for ELA and math, and 0.5 for science) and its IRT parameters. Each polytomously scored item appears several times in the OIB, once for each of its nonzero score points. Also, each page in the OIB corresponds to a score on the same scale. The item parameter estimates obtained after calibration were placed on a temporary standard-setting scale by applying the following linear transformation to the calibration θ scale:

$$\text{Standard-setting ability scale} = 50 \times \theta + 500$$

The same linear transformation was used to develop number-correct-to-scale-score tables and applied to student score distributions as the standard-setting performance impact data.

6.2.2. Procedures

The standard-setting workshop, led by Dr. Karla Egan and covering ELA and math grades 3–8 and 10, and science grades 5 and 8, took place in Topeka, Kansas, June 9–13, 2025. The goal of the event was to obtain cut-score recommendations from Kansas educators that differentiate four performance levels for the aforementioned subject areas and grades. To streamline the standard-setting process and promote consistency across similar grade levels, a mixture of grade-specific (single-grade) groupings and grade-span (bi-grade) groupings were implemented for the BSSP. Single-grade groupings were

implemented for ELA and math grade 10 and science grades 5 and 8. Bi-grade groupings for the KAP were as follows:

- Elementary: grades 3 and 4
- Intermediate: grades 5 and 6
- Middle school: grades 7 and 8

These groupings support alignment within each span, minimizing discrepancies in performance expectations, as students progress from one grade to the next within a span. The use of grade-span groupings also aids in the vertical articulation phase of the workshop, where panelists review and ensure consistency in cut scores across the entire grade sequence.

Grades 3–8 ELA and mathematics (bi-grade) panels were set up as a three-day workshop, which started in the morning of June 9 and ended in the afternoon of June 11. Grade 10 ELA and mathematics and grades 5 and 8 science (single-grade grouping) were set up as a two-day workshop, which started in the morning of June 12 and ended in the afternoon of June 13. During the standard-setting workshop, the panels recommended three scale-score cut points to delineate four performance levels for each grade-level assessment. The following sections describe the panelists, the PLDs, and the standard-setting procedure and outcomes for the standard-setting event.

Panelist Recruitment and Participation

Assessment & Technology Solutions (ATS) staff, in collaboration with KSDE, recruited and selected educators to serve as panelists for the standard-setting meeting. KSDE reviewed and finalized the panelist selection criteria, specifying that all middle and high school panelists should be licensed educators. ATS sought to fill the panels with educators who met the following criteria:

- They work in the grade level or band they represent.
- They work with students appropriate to the panel they're assigned to.
- They have experience in a variety of locations and contexts: rural and urban, large and small districts, varying socioeconomic statuses.
- They have a range of teaching experience levels.
- They represent all of the State Board of Education (SBOE) districts.
- They have experience teaching students with disabilities, English learners (ELs), and students from diverse backgrounds with regard to ethnicity, race, and gender.

Each discussion table within a panel included one seat for a school administrator (e.g., curriculum coordinator, testing coordinator). Classroom teachers and administrators from the same school district did not serve on the same panel.

Outside observers, such as TAC members and KSDE staff, observed portions of the standard-setting event. Specific agenda items were highlighted to support focused

observations, with limited interaction between observers and panelists to allow uninterrupted work.

Table 6-1 shows the panelist counts by table and grade. The recruiting goal for each bi-grade panel was 15 educators. The recruiting goal for each single-grade panel was 10 educators. Additional educators (universal panelists) who could serve in multiple grade panels were also recruited in case there were last-minute absences on the day of the workshop. These universal panelists were assigned to appropriate panels; therefore, the number of panelists was slightly higher than the original goals. A total of 141 educators participated in the workshop.

Table 6-1. Panelist Count by Table Assignment and Grade

Subject	Panel	Table 1	Table 2	Table 3	Total
ELA	Grades 3–4	6	5	5	16
	Grades 5–6	5	5	5	15
	Grades 7–8	6	6	5	17
	Grade 10	6	5	—	11
Mathematics	Grades 3–4	6	4	6	16
	Grades 5–6	5	6	5	16
	Grades 7–8	6	5	5	16
	Grade 10	5	5	—	10
Science	Grade 5	6	6	—	12
	Grade 8	6	5	—	11

As shown in Table 6-2, the majority of panelists were female, with most having personal or professional experience working with students from rural and suburban areas, and a few from urban settings. A significant number had over 10 years of experience in ELA and math, while some had between 5 and 10 years. Most panelists had experience working with students with disabilities or diverse learner populations, including those who are blind or visually impaired, who are deaf or hard of hearing, or who have cognitive, physical, or mobility disabilities. Many also had experience supporting ELs. The majority were classroom teachers, with a smaller number serving in administrative or support roles. In addition to those assigned to specific subject areas and grade levels, a few panelists demonstrated flexibility and contributed across multiple groups.

Table 6-2. Panel Composition

Subgroup	ELA				Math				Science		All Subjects Universal
	G3-4	G5-6	G7-8	G10	G3-4	G5-6	G7-8	G10	G5	G8	
Gender											
Female	15	16	14	7	14	15	14	8	8	10	6
Male	1	1	0	2	2	2	2	2	1	0	2
Nonbinary	0	1	0	0	0	1	0	0	1	0	0
Area experience											
Rural	17	14	14	8	14	16	18	9	8	9	10
Suburban	15	15	10	9	11	18	12	9	11	9	5
Urban	9	11	7	7	5	12	4	4	8	6	3
Teaching experience											
1-5 years	0	0	0	0	0	2	0	0	1	0	0
6-10 years	2	2	1	0	4	3	1	3	5	2	2
11+ years	14	16	13	9	12	13	15	7	4	8	6
Experience teaching SWDs											
Yes	16	16	13	9	17	16	16	14	13	13	12
No	0	0	1	0	1	2	1	0	1	0	1
Experience teaching ELs											
Yes	10	15	14	9	12	16	14	10	11	11	11
No	0	0	1	0	1	1	2	0	0	0	0
Role											
Teacher	14	15	12	9	13	16	14	10	9	10	7
Other	2	3	2	0	3	2	2	0	1	0	1

Note. SWD = student with disability; EL = English learner.

Performance Level Descriptors

Performance level descriptors (PLDs) played an integral role in guiding panelists' judgments during the standard-setting process. The following types of PLDs are used in the BSSP to provide clarity and coherence in defining performance standards:

- **Policy PLDs:** High-level descriptions of each performance level, reflecting policymakers' vision for student readiness. These descriptors are presented to panelists at the outset of the standard-setting process to establish the goals and rigor expected at each level. Table 6-3 shows the Kansas Policy PLDs.
- **Range PLDs:** Content-specific descriptors developed for each grade and subject area. These provide item writers and panelists with a clear understanding of the cognitive and content demands within each performance level.

The PLDs are designed to support consistent interpretation of the performance levels across assessment content areas, grades, and stakeholders. Policy PLDs are the same across grades and subjects for the Kansas Standards. The KAP assessment uses policy PLDs to report student performance on score reports and to define the general expectations for student performance with four levels. The four levels categorize student performance and describe what students likely know and can do relative to the academic content standards. During the BSSP process, panelists refer to policy PLDs to anchor their judgments, with the range PLDs refined as needed after the finalization of cut scores.

Table 6-3. Kansas Policy PLDs

Achievement Level	Policy Level Definitions
Level 1	A student at Level 1 shows a limited ability to demonstrate their knowledge and skills of (third, fourth, fifth, etc.) grade (ELA, Math, Science) standards.
Level 2	A student at Level 2 shows a basic ability to demonstrate their knowledge and skills of (third, fourth, fifth, etc.) grade (ELA, Math, Science) standards.
Level 3	A student at Level 3 shows a proficient ability to demonstrate their knowledge and skills of (third, fourth, fifth, etc.) grade (ELA, Math, Science) standards.
Level 4	A student at Level 4 shows an advanced ability to demonstrate their knowledge and skills of (third, fourth, fifth, etc.) grade (ELA, Math, Science) standards.

Note. ELA = English language arts.

Standard-Setting Procedures

The BSSP was conducted in a structured, multi-day workshop, beginning with an opening session and training to prepare panelists for each stage of the process. The workshop included three rounds of bookmark placement, followed by a final vertical articulation activity to ensure coherence across grades. Table 6-4 shows the daily agenda for the bi-grade panel.

Table 6-4. Agenda for Bi-grade Panels

Day 1	
Time	Activities
8:30 – 9:30 a.m.	Opening session Training Readiness survey General materials Standards
9:30 – 9:45 a.m.	Break / breakout rooms
9:45 a.m. – 12:00 p.m.	Lower grades (3, 5, 7) Threshold PLD discussions Round 1: Discuss OIB
12:00 – 12:30 p.m.	Lunch
12:30 – 5:00 p.m.	Continue Round 1
Day 2	
Time	Activities
8:30 – 9:00 a.m.	Round 2 training
9:00 – 11:00 a.m.	Round 2: Discuss cut-score placement
11:00 a.m. – 12:00 p.m.	Round 3: Show impact data Final cut-score recommendations
12:00 – 12:30 p.m.	Lunch
12:30 – 5:00 p.m.	Upper grades (4, 6, 8) Round 1
Day 3	
Time	Activities
8:30 – 10:30 a.m.	Round 2
10:30 a.m. – 12:00 p.m.	Round 3 Standard-setting evaluation
12:00 – 12:30 p.m.	Lunch
12:30 – 5:00 p.m.	Cross-grade articulation Articulation evaluation

Note. PLD = performance level descriptor; OIB = ordered item booklet.

Opening Session

The workshop began with an opening session in a large room, where panelists from all panels gathered for an introduction to the goals and importance of the standard-setting process. KSDE representatives provided a brief overview of the KAP, the history of the assessment standards, and the role of the Bookmark standard-setting procedure. This session emphasized the critical contribution of accurate performance standards to meaningful student assessment. Additionally, the policy PLDs were introduced to the panelists to set the foundational expectations for each performance level and the rigor required for cut scores.

Following the welcome from KSDE, lead facilitator Dr. Karla Egan guided panelists through the BSSP. This training included:

- Overview of the Bookmark process: An explanation of the BSSP methodology, including the purpose of each round, the use of bookmarks, and the role of impact data in round 3.
- Familiarization with materials: Panelists were introduced to the OIB and item map, which organize items by difficulty. They also learned how to reference the two types of PLDs (policy and range) during the Bookmark process.

After the training, a readiness survey assessed panelists' understanding and comfort level with the Bookmark process, allowing facilitators to provide additional guidance if needed. There were 89 (out of 97) and 37 (out of 44) panelists who indicated they were ready for the bi-grade panels and single-grade panels, respectively (no additional training needed nor questions unanswered). Dr. Egan visited the individual breakout room to provide support for panels that requested extra support. Appendix H | 2025 KAP Standard Setting – Opening Session Readiness Evaluation Results shows the results for the opening session survey.

Round 1: Initial Bookmark Placement

Panelists went to the breakout room for their grade or grade band after the opening session. Their work began by experiencing the assessment as a student would, familiarizing themselves with the test content and format. Panelists then discussed the range PLDs and identified skills and knowledge required for students who are minimally able to enter each performance category. The facilitator documented the skills and knowledge identified during the panelists' discussion. A detailed review of the OIB was followed. Once the study of the OIB was complete, panelists then gathered back in the large room.

Bookmark Training

Dr. Egan trained panelists from all grade and grade-band panels on the mechanics of setting a cut score. This training took about 45 minutes. Prior to placing bookmarks, panelists completed a round 1 readiness survey to determine their readiness to proceed with the first bookmark placement. There were 95 (out of 96) and 44 (out of 44) panelists who indicated they were ready to proceed after the Bookmark training for the bi-grade

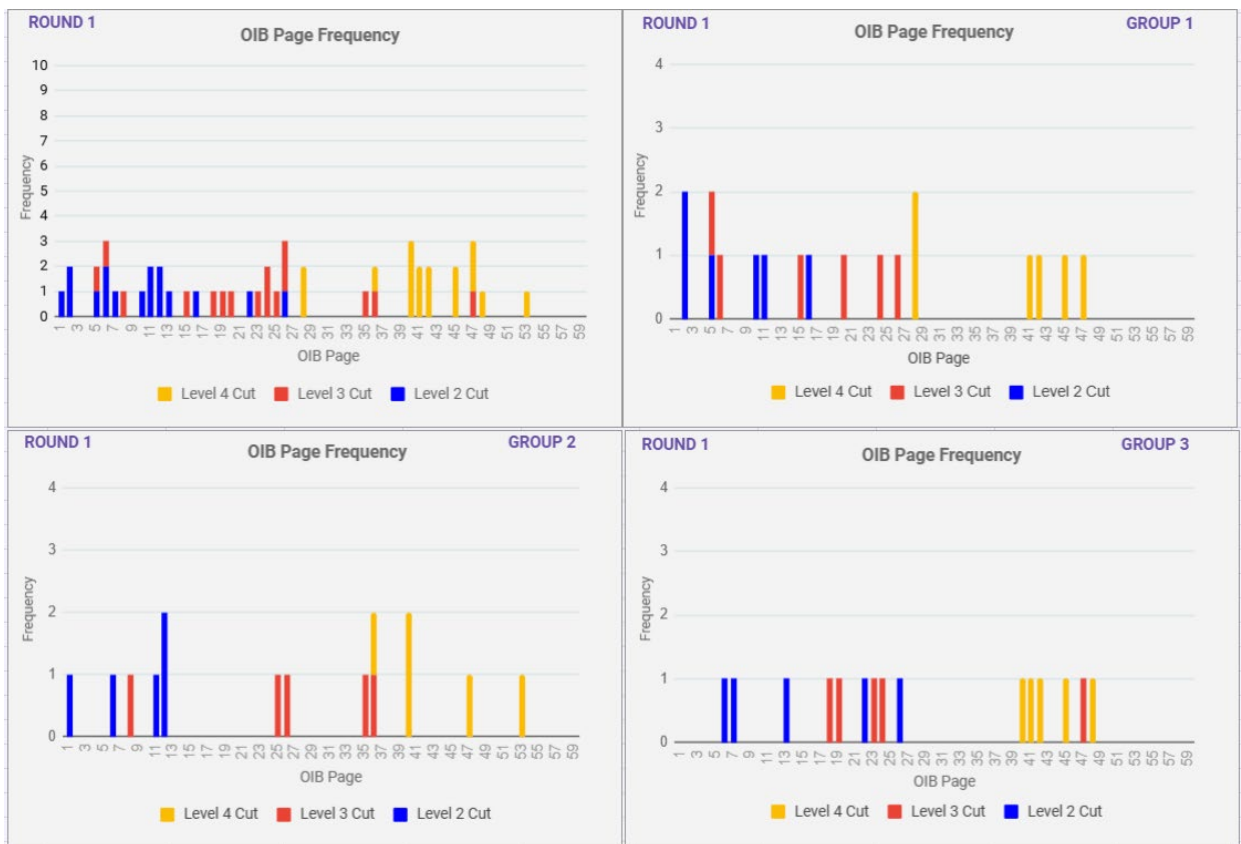
panels and single-grade panels, respectively. Dr. Egan provided additional training for panelists who wanted additional training or had additional questions. Appendix I | 2025 KAP Standard Setting – Round 1 Readiness Evaluation Results shows the results for the Bookmark readiness survey results.

When panelists indicated that they were prepared to make their judgments, they went back to the breakout room and independently reviewed the items in the OIB and placed their first bookmarks, beginning with the level 3 cut and proceeding to the level 2 and level 4 cuts.

Round 2: Table and Group Medians

During round 2, panelists reviewed the median round 1 cut-score recommendations for their individual table and for the overall group. They were also shown the distribution of individual ratings from their tables and overall group to understand the range of perspectives. Figure 6-1 shows an example of the bookmark page number rating distributions shown to panelists. Table leaders facilitated a discussion on items where there was significant variability in recommendations. Panelists then were asked to provide round 2 ratings of their bookmark placements, informed by both the median values and group discussions.

Figure 6-1. Panelist Rating Example



Round 3: Final Placement and Impact Data

In the final round, panelists were presented with the table medians, overall medians, and impact data, showing the percentage of students who would fall into each performance level based on the proposed cut scores. This impact data included 2024 operational impact data (that is not well articulated) and National Assessment of Educational Progress (NAEP) 2024 grade 4 and 8 ELA and math impact data. For grade 10 ELA and math panels, ACT college readiness performance in English and math were also presented. These data allowed panelists to assess the potential impact of their recommendations on student classification. After reviewing and discussing the data, panelists made final adjustments to their bookmark placements, and the cut scores recommended by the final panel were established based on the median values from this round. The impact data based on group median values were shared with panelists. Table 6-5, Table 6-6, and Table 6-7 show three rounds of ratings in the standard-setting scale-score metric. The ranges for each round of ratings became smaller by each round, showing panelists come to similar perspectives about where the cut scores should be.

Table 6-5. KAP ELA Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric)

Grade		Round 1			Round 2			Round 3		
		Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	Min	436.0	465.5	506.5	436.0	469.0	506.5	436.0	473.5	524.0
	Median	479.0	496.5	558.5	471.0	493.0	558.0	466.0	493.0	561.0
	Max	502.5	591.0	652.0	473.5	510.5	591.0	473.5	502.5	627.5
4	Min	417.0	476.0	520.5	417.0	476.0	524.5	444.0	492.5	532.5
	Median	444.0	486.0	531.5	451.0	486.0	531.0	452.0	493.0	564.0
	Max	473.0	492.5	563.5	469.5	492.5	532.5	469.5	512.5	580.5
5	Min	453.5	478.0	529.0	463.5	478.0	530.5	463.5	488.5	535.5
	Median	478.0	499.0	562.0	478.0	499.0	545.0	478.0	496.0	564.0
	Max	495.5	529.5	822.0	482.5	529.0	563.5	492.5	545.0	586.0
6	Min	437.5	479.5	507.5	440.5	479.5	534.5	440.5	491.0	535.5
	Median	462.5	502.0	550.0	462.0	502.0	547.0	463.0	502.0	552.0
	Max	502.0	550.0	575.0	479.5	534.5	571.0	484.0	535.0	571.0
7	Min	444.0	462.5	533.5	455.5	486.5	538.5	446.5	491.0	556.5
	Median	463.5	501.5	570.0	460.0	501.0	556.0	461.0	502.0	593.0
	Max	530.5	556.5	676.0	466.5	530.5	577.5	462.5	540.5	648.0
8	Min	461.0	483.5	530.5	470.0	488.0	526.0	470.0	502.0	566.5
	Median	477.5	509.5	558.0	475.0	509.0	563.0	472.5	523.5	570.0
	Max	487.0	526.0	584.5	483.5	519.5	568.0	483.5	535.5	584.5
10	Min	432.0	476.5	508.5	458.5	486.5	522.5	447.0	486.5	534.0
	Median	473.0	519.5	580.5	473.0	511.0	580.0	473.0	512.0	581.0
	Max	491.5	572.5	699.0	476.5	544.5	589.5	473.0	556.5	620.5

Table 6-6. KAP Math Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric)

Grade		Round 1			Round 2			Round 3		
		Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	Min	446.0	487.5	532.5	449.0	490.5	542.5	449.0	490.5	542.5
	Median	468.5	505.0	550.5	467.0	502.0	545.0	467.0	503.0	548.0
	Max	505.0	547.5	624.0	490.5	529.5	585.0	471.5	526.0	585.0
4	Min	446.5	469.5	512.0	467.5	482.5	512.5	467.5	482.5	523.5
	Median	475.0	492.5	533.5	475.0	494.0	533.0	470.0	495.0	540.0
	Max	481.5	528.0	549.0	480.5	506.0	542.5	480.5	511.0	543.0
5	Min	453.5	500.0	548.0	457.0	500.0	548.0	457.0	500.0	548.0
	Median	500.0	536.0	597.0	464.0	513.0	594.0	464.0	507.0	594.0
	Max	526.0	597.0	743.0	483.5	536.0	597.0	480.0	522.5	597.0
6	Min	444.5	467.0	530.0	460.5	470.5	555.5	460.5	487.0	558.0
	Median	463.5	496.5	574.5	463.0	481.0	570.0	464.0	501.0	570.0
	Max	479.0	530.0	1003.5	478.0	502.0	1003.5	479.0	503.0	594.0
7	Min	478.5	500.5	544.5	482.0	500.5	544.5	482.0	500.5	544.5
	Median	485.0	541.0	618.0	483.0	541.0	615.0	482.0	525.0	583.0
	Max	515.0	582.5	653.5	500.5	564.0	625.5	490.0	536.0	615.5
8	Min	475.5	489.5	517.5	475.5	505.0	521.5	475.5	507.5	543.0
	Median	485.0	517.5	550.0	485.0	517.0	549.0	483.0	518.0	549.0
	Max	505.0	574.0	679.0	497.5	518.5	574.0	487.0	517.5	568.0
10	Min	469.0	519.5	592.5	486.0	554.5	592.5	469.0	519.5	577.0
	Median	521.5	574.5	598.0	514.0	569.0	594.0	492.0	537.0	592.0
	Max	534.0	592.5	640.5	519.5	574.5	598.0	505.5	564.0	594.5

Table 6-7. KAP Science Bookmark Placement Ratings (in Standard-Setting Scale-Score Metric)

Grade		Round 1			Round 2			Round 3		
		Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
5	Min	385.0	462.0	548.0	456.5	487.5	564.0	456.5	487.5	564.0
	Median	465.0	533.5	603.5	462.0	519.0	592.0	462.0	509.5	571.0
	Max	526.0	584.5	701.0	467.0	526.0	592.5	473.5	521.5	603.5
8	Min	453.0	471.0	533.5	474.5	505.5	516.0	474.5	498.0	533.5
	Median	485.0	514.0	542.5	480.0	512.0	542.0	481.0	512.0	544.0
	Max	491.0	533.5	579.5	486.5	523.5	548.0	486.5	514.0	548.0

Vertical Articulation

After final bookmark placements, panelists engaged in a cross-grade vertical articulation activity. One panelist per table was selected to represent the whole panel during vertical articulation. There were two vertical articulation panels, one for ELA grades 3–8 and one for math grades 3–8. Each included nine panelists. The articulation meetings were held on the afternoon of the third day in the same room. Panelists not selected for articulation were free to observe the articulation process.

Dr. Egan presented the 2024 KAP student performance bar charts for all grades, the 2024 NAEP results for grades 4 and 8, and the impact data based on round 3 recommendations for grades 3–8. She then asked the articulation panels if the cut score and impact data made sense from one grade to another and for each subject. Each panel, facilitated by a facilitator, was asked to adjust cut scores and was shown impact data based on the suggested cut scores and what it would mean for the bookmark. This step allowed panelists to review the system of cut scores across grades, ensuring coherence in performance standards and impact data from grade 3 to grade 8 in ELA and mathematics.

A similar process took place for the single-grade panels in ELA and math grade 10, and in science grades 5 and 8. Vertical articulations were held for the single-grade panels in the afternoon of the last day of the workshop. The three articulation panels (ELA, math and science) were held in three separate rooms. Articulation results for ELA and math grades 3–8 from the bi-grade panels, 2024 KAP student performance for all grades, and round 3 recommendations were provided for panelists to review. Table 6-8 shows the panel recommendations after articulation.

Table 6-8. Articulation Recommendations

Subject	Grade	Level 2	Level 3	Level 4
ELA	3	463	492	568
	4	467	499	577
	5	473	509	589
	6	467	504	576
	7	464	504	591
	8	474	517	604
	10	470	512	576
Mathematics	3	465	502	542
	4	463	501	552
	5	464	498	570
	6	463	498	567
	7	469	500	576
	8	466	499	565
Science	10	471	508	592
	5	459	504	566
	8	478	511	544

Note. ELA = English language arts; the scores are expressed in a temporary standard-setting scale and do not reflect the final reporting scale.

Evaluation

At the end of the standard-setting meeting, panelists filled out the standard-setting evaluation survey. After the vertical articulation, participating panelists completed the articulation evaluation survey.

The standard-setting evaluation survey solicited feedback from panelists about the cut scores after completing all three rounds of bookmarking. Among the bi-grade panels, 94 out of 97 panelists responded, resulting in a response rate of 97%; among the single-grade panels, 43 out of 44 panelists responded, resulting in a response rate of 98%. In general, panelists in both bi-grade and single-grade panels indicated that they understand the standard-setting process and felt they could explain and support their recommendations. They also indicated that participating in the workshop increased their understanding of the Kansas assessments and that the workshop was a valuable professional development experience. See Table 6-9 for details.

Table 6-9. Results of the Standard-Setting Evaluation

Question	Bi-grade Panels		Single-Grade Panels	
	Strongly Agree	Agree	Strongly Agree	Agree
I understood how to place bookmarks in ordered item booklets.	79%	21%	79%	19%*
I felt that the Bookmark standard-setting procedure allowed me to recommend cut scores that reflected my thinking.	75%	25%	81%	14%*
I felt that the Bookmark standard-setting procedure was fair.	60%	40%	70%	28%*
I found the following resources helpful in placing bookmarks:				
My experience with students in grade level and subject	76%	25%	84%	16%
Performance level descriptors (PLDs)	72%	28%	74%	23%*
My experience of the Kansas Standards	69%	31%	70%	30%
My perceptions of the difficulties of the items	57%	42%	47%	51%*
I can explain and support the panel's following recommended cut scores against criticism:				
Proficient (level 3)	45%	54%*	61%	39%
Advanced (level 4)	46%	53%*	67%	33%
Basic (level 2)	48%	52%	63%	35%*
I understood that the purpose of this procedure was to set new cut scores for the 2025 KAP assessments.	86%	14%	91%	9%
Participating in the workshop increased my understanding of the Kansas assessments.	90%	10%	81%	19%
Overall, I believe that my opinions were considered and valued by my group.	79%	21%	88%	12%
Overall, I valued the workshop as a professional development experience.	87%	13%	84%	16%
The food and service at the facility met my expectations.	50%	42%*	58%	37%*
The workspace had accommodations appropriate to facilitate our work.	77%	22%*	84%	16%
The workshop was well organized.	81%	19%	86%	12%*

Note. KAP = Kansas Assessment Program; "*" indicates few negative responses were received for the corresponding questions.

The articulation evaluation survey was about panelists' perceptions of process of articulation. Among the bi-grade panels, 18 panelists responded, resulting in a response rate of 100%⁹; among the single-grade panels, 4 out of 6 panelists responded, resulting in a response rate of 67%. Due to the small number of respondents, results were pooled across all panels for reporting simplicity. Most panelists reported the session was well organized and the process was fair. They also reported that the articulation recommendations were reasonable, and they could explain and support the articulation recommendations. Table 6-10 shows the responses to the survey questions.

Table 6-10. Results of the Articulation Panel Evaluation Across All Panels

Question	Strongly Agree	Agree	Strongly Agree	Agree
The articulation session was well organized.	71%	25%	4%	0%
I had enough time to review cut scores for all grades recommended by the panels.	75%	17%	8%	0%
I felt that the articulation procedure was fair.	59%	33%	8%	0%
The PLD policy definitions were clearly communicated.	42%	17%	29%	12%
I used the PLDs to help me review my cut scores.	25%	17%	25%	33%
I can explain and support the panel's following recommended cut scores against criticism:				
Proficient (level 3)	54%	46%	0%	0%
Advanced (level 4)	58%	42%	0%	0%
Basic (level 2)	54%	42%	4%	0%
I feel the recommended cut scores that resulted from this process are reasonable.	54%	42%	4%	0%
Overall, I believe that my opinions were considered and valued by my group.	76%	12%	12%	0%
Overall, I valued the workshop as a professional development experience.	84%	12%	4%	0%

Note. PLD = performance level descriptor.

⁹ The evaluation surveys were anonymous. Some of the articulation survey results might be provided by the observers instead of the articulation panelists.

6.3. Challenging and Aligned Academic Achievement Standards

6.3.1. Development of Grade-Specific Performance Level Descriptors

Educators set the KAP’s academic achievement standards to align with the state content standards (i.e., Kansas Standards). First, ATS content experts worked alongside KSDE staff to define the scope of work for the different subject areas in developing the PLDs.

The overarching goal was to have standard-level PLDs with vertical coherence and standard alignment. In ELA, the task was more revision than generation because ELA already had existing PLDs articulated at the standard level. There were only a handful of standards for which new PLDs needed to be drafted in ELA due to blueprint changes since the last time they had been developed. For ELA, the task mainly was to update the terminology to reflect the terminology in the new standards, tighten the language in the PLDs to create meaningful distinctions between the levels, and ensure that the full set of assessed standards was reflected in the PLDs.

For math, the existing PLDs were articulated at the cluster level, so the PLDs needed to be “brought down” to the standard level. This was a development-heavy task because essentially new PLDs needed to be developed for every standard at each grade.

For science, it was also a development more than a revision effort to create PLDs for each assessed standard in grades 5, 8, and 11.

The development process involved multiple levels of review. Internally, each PLD in each subject area had a content review and an editorial review. Then, KSDE reviewed the PLDs. After the KSDE review stage came an educator review. Each PLD was reviewed by at least two external educator reviewers.

The educator review was an unpaid professional opportunity. Recruitment for the educator review was a joint KSDE and ATS effort, and KSDE had final approvals on educator reviewers. We strove for a broad range of educator representation across the SBOEs in Kansas.

Based on their experience and expertise, reviewers would review two contiguous grades (except in science), answer two primary questions, offer feedback on a third inquiry, and have the opportunity to “write in” any feedback. The two primary questions were: “Do the PLDs have vertical coherence? Do the PLDs have horizontal coherence?” The third inquiry was this: “In your opinion, do the PLDs seem to be showcasing skills that are low in difficulty, medium in difficulty, or high in difficulty?”

Post-educator review, the review comments and suggested revisions/edits were shared with KSDE, who approved any changes. The changes were minimal across all subject areas because the reviewers widely agreed that the PLDs had vertical and horizontal coherence. Once the educator review feedback had been applied, the PLDs were ready for finalization for standard setting. Finalization involved editorial review and creating PDFs.

The grade-specific PLDs are range PLDs that were developed using professional judgment. Items and item-level data were not considered during the development of the PLDs.

The priorities for these PLDs were articulated by the lead psychometrician.

The PLD goals are the following:

- **Vertical coherence:** The grade-specific PLDs will clarify how expectations change as grade levels increase, so there is vertical coherence across grades.
- **Cognitive process:** The grade-specific PLDs should reflect the cognitive processes required for specific content areas and describe what students know and can do at each performance level.
- **Adherence to content standards:** The grade-specific PLDs should closely align with the content standards and address at least the cluster/target level.

The range PLDs helped guide educator thinking during standard setting itself around the threshold skills.

6.3.2. Academic Performance Standards

The Kansas SBOE approved the Kansas assessment scale and cut scores for the 2025 revised state assessments in ELA, mathematics, and science in August 2025. Table 6-11 shows the score scale and cut scores.

Table 6-11. Kansas Assessment Scale and Cut Scores for the 2025 Revised State Assessments

Subject	Grade	LOSS	Level 2 Cut	Level 3 Cut	Level 4 Cut	HOSS
ELA	3	400	510	540	609	700
	4	400	510	540	607	700
	5	400	510	540	600	700
	6	400	510	540	593	700
	7	400	510	540	600	700
	8	400	510	540	594	700
	10	400	510	540	588	700
Mathematics	3	400	510	540	570	700
	4	400	510	540	579	700
	5	400	510	540	599	700
	6	400	510	540	596	700
	7	400	510	540	603	700
	8	400	510	540	594	700
Science	5	400	510	540	582	700
	8	400	510	540	569	700

Note. ELA = English language arts; LOSS = lowest obtainable scale score; HOSS = highest obtainable scale score.

6.4. Reporting

For each tested grade and subject, the KAP assessment provides separate score reports to students, schools, and districts. These reports include students' overall and subscore performances. Score reports present the results using various graphs, colors, and symbols so they are easy to read. To assist readers in interpreting the information in the reports, descriptions of what students should be able to do in each subject area are presented with the results. As stated by Petersen, Kolen, and Hoover (1989), providing score interpretations in score reports can minimize misinterpretations and unwarranted inferences. Helping readers understand the meaning of the statistics is as important as reporting the values.

Although these reports are intended for different groups (e.g., students, schools, and districts), the content of these reports is uniform. Presentation and text are adjusted according to group, but the symbols and interpretation of those symbols are consistent across reports. The uniform design eases educators' burden of review and helps them explain score reports to parents.

6.4.1. Student Reports

The individual student report is made available for each student taking the KAP

assessment. Each content area has its own report. A student may receive up to three KAP individual student reports for ELA, math, and science. The [KAP Summative Parent Guide to Reports](#) includes a sample of an individual student report and narratives of information presented in the student report.

There are eight sections of the report. The first five sections are on the front page. The front page of the report provides information on the student's overall performance for the content area. Subscore information is reported on the back page of the report. It indicates strengths and weaknesses in different reporting categories.

- Section 1: Student identifier and a brief description of the content area reported
- Section 2: State-, district-, and school-level average scale scores and number of students per level
- Section 3: Student's scale score and performance level
- Section 4: Performance level descriptors and the cut scores separating each performance level
- Section 5: Conversation starters for parents to talk to their student or the teacher
- Section 6: Brief description of reporting categories (subscores)
- Section 7: Student performance in each reporting category
- Section 8: Links to resources to improve students' performance and Lexile/Quantile measure

6.4.2. School and District Reports

While student reports focus on individual student performance, school and district reports focus on group-level performances. Information provided in the school and district reports aggregates student performances at the given performance level (Level 1 through Level 4). The [KAP Summative Educator Guide to Reports](#) includes information about how to interpret the school and district reports.

School reports provide summary information of a subject by grades. On the first page, bar graphs indicate a school's average scaled scores of each grade, along with scores of the school's district and state overall performance. The report includes district and state average scaled scores as a reference so that schools can interpret their standing. Standard errors of the mean are given at the bottom of the first page to help educators take into account score variability when making inferences. The second page shows the percentage of students in each of the four performance levels; again, the school report provides district and state results for reference. The bar graphs use four different shades of blue to represent the different performance levels, allowing readers to distinguish performance-level outcomes instantly. The next section of the school report presents the school's aggregated performance for different reporting categories. Percentages of students receiving ratings of near or above the state average were reported for each reporting category.

District reports use the same layout and provide the same information as school reports with aggregation at the district level; however, only state data are provided as the reference group.

When group case counts are very small, individual students may be identified through demographic information, even on roll-up summary reports. For a school or district with fewer than 10 students, the school or district report is suppressed for KAP.

6.4.3. Reporting Timeline

The KAP testing window for brick-and-mortar schools ended on April 25, 2025. The KAP testing window for virtual schools ended on May 9, 2025. Achievement and Assessment Institute (AAI) and KSDE staff presented cut scores to the Kansas SBOE after the June 2025 standard-setting meeting. The Kansas SBOE approved the cut scores on August 12, 2025. KSDE reviewed and approved the reports prior to report release. Access to reports was provided to districts and parents on September 9, 2025.

6.4.4. Interpretive Guides

To help educators and parents interpret KAP results, the [KAP Summative Educator Guide to Reports](#) and the [KAP Summative Parent Guide to Reports](#) (including a [Spanish version](#)) are available on the KAP website so that educators and parents can access them easily. Both guides include an overview of test purposes, content, and format; descriptions of the KAP scoring process; suggestions for how to use test scores and improve KAP scores; a sample of the score report; and explanations of different information presented on the score reports.

References

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–472). Reading, MA: Addison Wesley.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications Ltd.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29.
<https://doi.org/10.18637/jss.v048.i06>
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360.
- Hohensinn, C., Kubinger, K. D., Reif, M., Schleicher, E., & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation*, 17(6), 497–509.
<https://doi.org/10.1080/13803611.2011.632668>
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rate using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329–349.
https://doi.org/10.1207/S15324818AME1404_2
- Kan. Stat. Ann. §72-5170 (2024).
https://www.ksrevisor.org/statutes/chapters/ch72/072_051_0070.html

- Kansas State Department of Education. (2021). Navigating change: Kansas' guide to learning and school safety operations.
https://www.ksde.gov/Portals/0/Communications/Navigating%20Change/Navigating_Change.pdf
- Li, F., Cohen, A., & Shen, L. (2012). Investigating the effect of item position in computer-based tests. *Journal of Educational Measurement*, 49(4), 362–379.
<http://www.jstor.org/stable/23353876>
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179–197.
- Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262). New York, NY: American Council on Education/Macmillan.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes: The art of scientific computing* (3rd ed.). University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.
- Thissen, D., & Wainer, H. (2001). *Test Scoring*. Lawrence Erlbaum Associates.
- Webb, N. L. (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Research Monograph No. 6). National Institute for Science Education Research and Council of Chief State School Officers. ERIC Document Reproduction Service.
<https://files.eric.ed.gov/fulltext/ED414305.pdf>

Appendix A | Test Blueprints

Table A-1. Grade 3 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes	20–30%
Language in Writing	
Reading: Literary	
Key Ideas & Details	35–45%
Craft & Structure	
Integration of Knowledge & Ideas	
Language in Reading	
Reading: Informational	
Key Ideas & Details	30–40%
Craft & Structure	
Integration of Knowledge & Ideas	
Language in Reading	

Table A-2. Grade 4 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes	20–30%
Language in Writing	
Reading: Literary	
Key Ideas & Details	35–45%
Craft & Structure	
Integration of Knowledge & Ideas	
Language in Reading	
Reading: Informational	
Key Ideas & Details	30–40%
Craft & Structure	
Integration of Knowledge & Ideas	
Language in Reading	

Table A-3. Grade 5 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes Language in Writing	20–30%
Reading: Literary	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	35–45%
Reading: Informational	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	35–45%

Table A-4. Grade 6 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes Language in Writing	15–25%
Reading: Literary	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	35–45%
Reading: Informational	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	40–50%

Table A-5. Grade 7 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes Language in Writing	15–25%
Reading: Literary	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	35–45%
Reading: Informational	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	40–50%

Table A-6. Grade 8 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes Language in Writing	20–30%
Reading: Literary	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	32–42%
Reading: Informational	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	37–47%

Table A-7. Grade 10 English Language Arts (ELA) Summative Assessment Blueprint

Domain/Subdomain	Percentage of Assessment
Writing	
Text Types & Purposes Language in Writing	20–25%
Reading: Literary	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	30–40%
Reading: Informational	
Key Ideas & Details Craft & Structure Integration of Knowledge & Ideas Language in Reading	35–50%

Table A-8. Grade 3 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Operations & Algebraic Thinking	20–35%
Number & Operations in Base 10	13–18%
Number & Operations - Fractions	13–18%
Measurement & Data	35–50%
Geometry	3–8%

Table A-9. Grade 4 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Operations & Algebraic Thinking	18–28%
Number & Operations in Base 10	20–30%
Number & Operations - Fractions	35–50%
Measurement & Data	8–18%
Geometry	3–13%

Table A-10. Grade 5 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Operations & Algebraic Thinking	3–8%
Number & Operations in Base 10	25–35%
Number & Operations - Fractions	38–48%
Measurement & Data	15–25%
Geometry	5–15%

Table A-11. Grade 6 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Ratios & Proportional Relationships	13–18%
The Number System	30–45%
Expressions & Equations	28–38%
Geometry	5–10%
Statistics & Probability	12–22%

Table A-12. Grade 7 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Ratios & Proportional Relationships	13–18%
The Number System	18–23%
Expressions & Equations	15–25%
Geometry	18–28%
Statistics & Probability	22–37%

Table A-13. Grade 8 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
The Number System	3–8%
Expressions & Equations	25–38%
Functions	22–32%
Geometry	25–35%
Statistics & Probability	12–17%

Table A-14. Grade 10 Math Summative Assessment Blueprint

Domain	Percentage of Assessment
Number & Quantity	0–5%
Algebra	18–28%
Functions	20–30%
Geometry	40–50%
Statistics & Probability	12–22%

Table A-15. Grade 5 Science Summative Assessment Blueprint

Domain	Percentage of Assessment
Physical Science	35–50%
Life Science	20–35%
Earth & Space Science	25–40%

Table A-16. Grade 8 Science Summative Assessment Blueprint

Domain	Percentage of Assessment
Physical Science	35–50%
Life Science	30–45%
Earth & Space Science	20–30%

Table A-17. Grade 11 Science Summative Assessment Blueprint

Domain	Percentage of Assessment
Physical Science	45–60%
Life Science	25–40%
Earth & Space Science	15–35%

Appendix B | Word Count and Lexile Range

Table B-1. Word Count and Lexile Range by Grade

Grade	Word Count	Lexile
3	500–750	420–760
4	550–800	740–950
5	600–800	830–1010
6	700–900	925–1165
7	750–950	970–1170
8	850–1,000	1010–1185
10	950–1,100	1050–1335

Appendix C | Demographic Information for Passage Reviewers

Table C-1. Demographic Information of Grade 3–5 Content Passage Review Panelists

Characteristic	Summer 2021	Winter 2022	%		Summer 2023
			Summer 2022	Winter 2023	
Sex					
Female	100	86	100	100	83
Male	0	14	0	0	17
Race					
White	85	57	100	100	83
Black	14	0	0	0	0
Asian	0	14	0	0	0
Native American	0	14	0	0	0
Other	0	14	0	0	17
Ethnicity					
Hispanic	0	0	0	0	0
Non-Hispanic	100	100	100	100	83
Role					
Classroom teacher	71	71	80	25	83
District staff	14	14	0	25	0
Other	14	14	20	50	17
SBOE district					
1	14	0	0	25	17
2	0	0	0	0	17
3	0	0	0	50	0
4	0	28	0	0	0
5	0	14	0	0	0
6	43	28	40	0	0
7	14	0	20	0	0
8	14	0	0	0	17
9	14	14	40	0	17
10	0	0	0	0	0
Multiple	0	0	0	0	33
Years of experience					
< 3	0	0	0	0	0
3–5	0	14	0	0	0
6–9	28	14	20	25	33
≥ 10	71	71	80	75	67

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table C-2. Demographic Information of Grade 6–8 Content Passage Review Panelists

Characteristic	Summer 2021	Winter 2022	%	
			Winter 2023	Summer 2023
Sex				
Female	100	100	100	100
Male	0	0	0	0
Race				
White	83	100	67	100
Black	0	0	0	0
Asian	0	0	0	0
Native American	0	0	0	0
Other	0	0	33	0
Ethnicity				
Hispanic	0	0	17	0
Non-Hispanic	100	100	83	100
Role				
Classroom teacher	83	88	83	83
District staff	0	12	0	0
Other	16	0	17	17
SBOE district				
1	16	13	0	33
2	16	13	17	0
3	0	0	0	0
4	0	0	17	0
5	16	25	17	17
6	16	13	0	17
7	0	0	17	17
8	0	13	17	0
9	16	25	17	17
10	16	0	0	0
Years of experience				
< 3	0	0	0	0
3–5	0	0	0	0
6–9	16	12	0	0
≥ 10	83	88	100	100

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table C-3. Demographic Information of Grade 10 Content Passage Review Panelists

Characteristic	Summer 2021	% Winter	
		2022	2023
Sex			
Female	100	86	67
Male	0	14	33
Race			
White	100	100	100
Black	0	0	0
Asian	0	0	0
Native American	0	0	0
Other	0	0	0
Ethnicity			
Hispanic	14	0	0
Non-Hispanic	86	100	100
Role			
Classroom teacher	86	86	67
District staff	14	14	0
Other	0	0	33
SBOE district			
1	0	0	17
2	14	0	0
3	14	29	33
4	14	29	17
5	28	14	17
6	0	0	0
7	14	0	17
8	0	0	0
9	0	0	0
10	14	29	0
Years of experience			
< 3	0	0	0
3–5	0	0	0
6–9	14	14	0
≥ 10	86	86	100

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table C-4. Demographic Information of Grade 3–5 Fairness Passage Review Panelists

Characteristic	Summer 2021	Winter 2022	% Summer 2022		
			Summer 2022	Winter 2023	Summer 2023
Sex					
Female	100	86	83	100	100
Male	0	14	16	0	0
Race					
White	66	100	83	67	86
Black	16	0	0	0	0
Asian	0	0	16	0	0
Native American	0	0	0	17	0
Other	16	0	0	17	14
Ethnicity					
Hispanic	16	29	0	17	14
Non-Hispanic	84	71	100	83	86
Role					
Classroom teacher	33	71	66	66	71
District staff	50	29	16	16	0
Other	17	0	16	16	29
SBOE district					
1	16	14	0	0	14
2	0	14	0	0	0
3	0	0	16	0	0
4	0	0	0	0	0
5	16	14	16	33	0
6	16	14	0	17	0
7	0	29	0	17	0
8	0	14	0	0	0
9	0	0	33	17	14
10	0	0	16	0	0
Multiple	0	0	0	0	57
Years of experience					
< 3	0	0	0	0	14
3–5	0	0	16	17	0
6–9	0	29	16	0	14
≥ 10	66	71	66	83	71

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table C-5. Demographic Information of Grade 6–8 Fairness Passage Review Panelists

Characteristic	Summer 2021	Winter 2022	%	
			Winter 2023	Summer 2023
Sex				
Female	100	60	100	83
Male	0	20	0	17
Race				
White	75	60	87	83
Black	12	20	13	17
Asian	0	0	0	0
Native American	0	0	0	0
Other	0	0	0	0
Ethnicity				
Hispanic	12	0	0	0
Non-Hispanic	62	80	100	100
Role				
Classroom teacher	50	80	63	100
District staff	50	0	13	0
Other	0	0	25	0
SBOE district				
1	25	0	13	0
2	0	20	0	0
3	0	0	0	0
4	12	40	25	0
5	25	20	13	17
6	12	0	0	17
7	12	0	25	0
8	0	0	13	0
9	0	0	0	17
10	0	0	13	0
Multiple	0	0	0	33
Years of experience				
< 3	0	0	0	0
3–5	0	20	0	0
6–9	0	0	0	17
≥ 10	62	60	100	83

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table C-6. Demographic Information of Grade 10 Fairness Passage Review Panelists

Characteristic	%		
	Summer 2021	Winter 2022	Winter 2023
Sex			
Female	57	67	67
Male	42	33	33
Race			
White	85	100	83
Black	0	0	0
Asian	0	0	0
Native American	0	0	17
Other	0	0	0
Ethnicity			
Hispanic	0	17	0
Non-Hispanic	100	83	100
Role			
Classroom teacher	70	83	67
District staff	14	17	17
Other	14	0	17
SBOE district			
1	0	0	0
2	0	0	17
3	14	17	0
4	14	0	17
5	0	33	0
6	14	0	0
7	0	0	0
8	0	17	33
9	28	0	33
10	0	33	0
Years of experience			
< 3	0	0	0
3–5	14	17	0
6–9	0	0	0
≥ 10	28	83	100

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Appendix D | Demographic Information for External Item Reviewers

Table D-1. Demographic Information of English Language Arts Content Item Review Panelists

Characteristic	%								
	Grades 3–5			Grades 6–8			Grade 10		
	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Winter 2023	Summer 2023
Sex									
Female	100	100	83	100	100	100	100	66	88
Male	0	0	17	0	0	0	0	33	12
Race									
White	86	100	83	83	100	100	100	100	100
Black	14	0	0	0	0	0	0	0	0
Asian	0	0	0	0	0	0	0	0	0
Native American	0	0	0	0	0	0	0	0	0
Other	0	0	17	17	0	0	0	0	0
Ethnicity									
Hispanic	0	0	0	0	0	0	14	0	12
Non-Hispanic	100	100	83	100	100	100	86	83	88
Role									
Classroom teacher	86	80	83	83	75	83	86	66	88
District staff	0	0	0	0	25	0	14	16	0
Other	14	20	17	17	0	17	0	0	12
SBOE district									
1	14	0	17	17	25	33	0	16	0
2	0	0	17	17	25	0	14	0	0
3	0	0	0	0	0	0	14	33	0
4	0	0	0	0	25	0	14	16	0
5	0	0	0	17	0	17	29	16	0
6	43	40	0	17	0	17	0	0	0
7	14	20	0	0	0	17	14	16	0

Characteristic	%								
	Grades 3–5			Grades 6–8			Grade 10		
	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Winter 2023	Summer 2023
8	14	0	17	0	0	0	0	0	0
9	14	40	17	17	25	17	0	0	25
10	0	0	0	17	0	0	14	0	38
Multiple	0	0	33	0	0	0	0	0	38
Years of experience									
< 3	0	0	0	0	0	0	0	0	0
3–5	0	20	0	0	0	0	0	0	12
6–9	29	0	33	17	50	0	14	0	0
≥ 10	71	80	67	83	50	100	86	100	88

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table D-2. Demographic Information of English Language Arts Fairness Item Review Panelists

Characteristic	%								
	Grades 3–5			Grades 6–8			Grade 10		
	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Winter 2023	Summer 2023
Sex									
Female	100	83	100	100	80	83	57	57	100
Male	0	17	0	0	20	17	43	28	0
Race									
White	66	83	86	88	100	83	86	71	50
Black	17	0	0	12	0	17	0	0	0
Asian	0	17	0	0	0	0	0	0	0
Native American	0	0	0	0	0	0	0	14	0
Other	17	0	14	0	0	0	14	0	33
Ethnicity									
Hispanic	17	0	14	12	0	0	0	0	0
Non-Hispanic	83	100	86	63	100	100	100	85	67
Role									
Classroom teacher	33	67	71	63	80	100	71	71	67
District staff	50	0	0	0	0	0	0	14	0
Other	17	33	29	38	20	0	29	0	17
SBOE district									
1	17	0	14	25	0	0	0	0	0
2	0	0	0	0	0	0	0	14	0
3	0	17	0	0	0	0	14	0	0
4	0	0	0	12	40	0	14	14	0
5	17	17	0	25	20	17	0	0	33
6	17	0	0	12	0	17	14	0	0
7	0	0	0	12	20	0	0	0	0
8	0	0	0	0	0	0	0	28	0
9	17	33	14	0	0	17	29	28	0

Characteristic	%								
	Grades 3–5			Grades 6–8			Grade 10		
	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Summer 2022	Summer 2023	Summer 2021	Winter 2023	Summer 2023
10	0	17	0	0	20	0	0	0	0
Multiple	0	0	57	0	0	33	0	0	50
Years of experience									
< 3	0	0	14	0	0	0	0	0	17
3–5	0	17	0	0	0	0	14	0	0
6–9	17	17	14	0	0	17	0	0	0
≥ 10	83	67	71	75	100	83	29	85	66

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table D-3. Demographic Information of Mathematics Content Item Review Panelists

Characteristic	Grades 3–5		Grades 6–8		Grade 10	
	Summer 2021	Summer 2022	Summer 2021	Summer 2022	Summer 2021	Summer 2022
Sex						
Female	100	100	83	33	71	60
Male	0	0	17	66	14	20
Race						
White	83	66	83	100	86	80
Black	0	0	0	0	0	0
Asian	0	0	0	0	0	20
Native American	17	33	0	0	0	0
Other	0	0	17	0	0	0
Ethnicity						
Hispanic	17	0	17	0	0	0
Non-Hispanic	83	100	83	100	86	100
Role						
Classroom teacher	83	100	83	67	100	100
District staff	0	0	0	33	0	0
Other	17	0	17	0	0	0
SBOE district						
1	33	0	17	0	0	20
2	0	0	0	33	0	0
3	17	0	33	0	14	20
4	17	33	0	0	14	0
5	0	0	0	0	0	0
6	17	0	0	33	0	0
7	0	33	17	0	29	0
8	17	0	0	0	14	0
9	0	33	17	33	29	40
10	0	0	17	0	0	20
Years of experience						
< 3	17	0	0	0	0	0
3–5	17	0	17	0	0	0
6–9	17	33	17	0	0	0
≥ 10	50	66	67	100	100	100

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table D-4. Demographic Information of Mathematics Fairness Item Review Panelists

Characteristic	Grades 3–5		Grades 6–8		Grade 10	
	Summer 2021	Summer 2022	Summer 2021	Summer 2022	Summer 2021	Summer 2022
Sex						
Female	67	100	100	33	85	66
Male	0	0	0	66	14	0
Race						
White	67	100	83	100	71	100
Black	0	0	0	0	0	0
Asian	0	0	0	0	0	0
Native American	0	0	0	0	0	0
Other	0	0	0	0	0	0
Ethnicity						
Hispanic	0	0	0	0	0	0
Non-Hispanic	66	100	83	100	85	100
Role						
Classroom teacher	50	100	66	100	85	100
District staff	16	0	16	0	0	0
Other	0	0	16	0	14	0
SBOE district						
1	0	0	0	0	0	33
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	16	0	28	0
5	0	0	0	0	0	33
6	33	33	16	0	0	0
7	16	33	16	0	14	0
8	0	0	0	0	0	0
9	0	0	33	66	14	33
10	0	33	0	33	14	0
Years of experience						
< 3	0	0	0	0	0	0
3–5	16	33	16	33	42	0
6–9	0	0	0	0	0	0
≥ 10	50	66	50	66	28	100

Note. SBOE = State Board of Education. Two out of six participants for the summer 2021 panel for grades 3–5 did not provide any demographic information. The summary is based on the remaining four participants. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table D-5. Demographic Information of Science Content Item Review Panelists

Characteristic	Grade 5		Grade 8		Grade 11
	Summer 2021	Summer 2022	Summer 2021	Summer 2022	Summer 2023
%					
Sex					
Female	67	100	75	66	100
Male	33	0	25	33	0
Race					
White	83	100	100	100	100
Black	0	0	0	0	0
Asian	0	0	0	0	0
Native American	0	0	0	0	0
Other	17	0	0	0	0
Ethnicity					
Hispanic	33	0	0	0	0
Non-Hispanic	67	100	100	100	100
Role					
Classroom teacher	83	75	100	66	100
District staff	0	0	0	33	0
Other	17	25	0	0	0
SBOE district					
1	17	0	25	33	17
2	17	0	0	0	0
3	0	50	0	33	0
4	0	0	25	0	17
5	33	0	0	0	0
6	17	25	0	33	0
7	0	0	0	0	0
8	17	0	0	0	17
9	0	0	0	0	0
10	0	0	50	0	0
Multiple	0	0	0	0	50
Years of experience					
< 3	0	0	0	0	0
3–5	17	0	0	0	0
6–9	33	25	0	0	17
≥ 10	50	75	100	100	83

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table D-6. Demographic Information of Science Fairness Item Review Panelists

Characteristic	Grade 5		Grade 8		Grade 11
	Summer 2021	Summer 2022	Summer 2021	Summer 2022	Summer 2023
Sex					
Female	100	75	50	50	83
Male	0	25	16	50	17
Race					
White	33	100	66	100	83
Black	0	0	0	0	0
Asian	0	0	0	0	0
Native American	0	0	0	0	0
Other	66	0	0	0	17
Ethnicity					
Hispanic	0	0	0	0	17
Non-Hispanic	100	100	66	100	83
Role					
Classroom teacher	66	100	50	100	100
District staff	0	0	16	0	0
Other	33	0	0	0	0
SBOE district					
1	0	0	16	50	0
2	0	0	0	0	0
3	33	0	33	0	0
4	0	0	16	25	0
5	0	75	16	0	50
6	33	0	0	0	0
7	0	25	0	0	0
8	0	0	0	25	17
9	0	0	0	0	0
10	33	0	0	0	0
Multiple	0	0	0	0	17
Years of experience					
< 3	0	0	0	0	0
3–5	33	0	0	0	17
6–9	0	0	0	50	50
≥ 10	33	50	16	50	33

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Appendix E | Item Statistics Flagging Criteria

Table E-1. Item Statistics Flagging Criteria

Statistic	Criteria
Omit	Omit correlation > .1 Omit percentage > .05
Differential item functioning	Gender R^2 change > 0.035 Race R^2 change > 0.035 Ethnicity R^2 change > 0.035 English learner R^2 change > 0.035
Item-total correlation	Item-total correlation \leq .25
p -value	p -value < 0.2 p -value > 0.9
Distractors for selecting-key items	Correlation of distractors > -0.05 Percentage of selecting distractor > Percentage of selecting keyed response

Appendix F | Demographic Information for Data Reviewers

Table F-1. Demographic Information of English Language Arts Data Review Panelists

Characteristic	Grades 3–5		%		Grade 10 Summer 2022
	Summer 2022	Summer 2023	Grades 6–8 Summer 2022	Grades 6–8 Summer 2023	
Sex					
Female	91	100	89	66	87
Male	9	0	11	33	13
Race					
White	91	100	100	100	100
Black	0	0	0	0	0
Asian	9	0	0	0	0
Native American	0	0	0	0	0
Other	0	0	0	0	0
Ethnicity					
Hispanic	0	0	0	17	0
Non-Hispanic	100	100	100	83	100
Role					
Classroom teacher	73	67	78	66	100
District staff	9	0	11	17	0
Other	9	33	11	17	0
SBOE district					
1	0	0	11	0	25
2	0	0	11	0	0
3	9	0	0	0	13
4	0	17	33	0	0
5	9	0	11	33	25
6	18	17	0	0	25
7	9	0	11	0	0
8	0	0	0	0	0
9	36	17	11	0	13
10	9	0	11	0	0
Multiple	0	50	0	50	0
Years of experience					
< 3	0	17	0	0	0
3–5	18	0	0	0	0
6–9	9	0	22	17	13
≥ 10	73	83	78	83	87

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table F-2. Demographic Information of Mathematics Data Review Panelists

Characteristic	Grades 3–5		Grades 6–8		Grade 10	
	Summer 2022	Summer 2023	Summer 2022	Summer 2023	Summer 2022	Summer 2023
Sex						
Female	100	83	33	66	63	83
Male	0	17	66	33	13	17
Race						
White	83	83	100	83	88	100
Black	0	0	0	17	0	0
Asian	0	0	0	0	13	0
Native American	17	17	0	0	0	0
Other	0	0	0	0	0	0
Ethnicity						
Hispanic	0	0	0	0	0	0
Non-Hispanic	100	100	100	100	100	100
Role						
Classroom teacher	100	83	83	83	100	100
District staff	0	17	17	0	0	0
Other	0	0	0	17	0	0
SBOE district						
1	0	17	0	17	25	17
2	0	0	17	0	0	17
3	0	0	0	0	13	33
4	17	0	0	0	0	0
5	0	0	0	0	13	33
6	17	0	17	0	0	0
7	33	0	0	0	0	0
8	0	17	0	0	0	0
9	17	17	50	17	38	0
10	17	0	17	17	13	0
Multiple	0	50	0	50	0	0
Years of experience						
< 3	0	0	0	0	0	50
3–5	17	50	17	17	0	17
6–9	17	17	0	17	0	0
≥ 10	66	33	83	66	100	33

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Table F-3. Demographic Information of Science Data Review Panelists

Characteristic	%			
	Grade 5		Grade 8	
	Summer 2022	Summer 2023	Summer 2022	Summer 2023
Sex				
Female	87	100	57	100
Male	13	0	43	0
Race				
White	100	100	100	100
Black	0	0	0	0
Asian	0	0	0	0
Native American	0	0	0	0
Other	0	0	0	0
Ethnicity				
Hispanic	0	0	0	0
Non-Hispanic	100	100	100	100
Role				
Classroom teacher	87	100	86	100
District staff	0	0	14	0
Other	13	0	0	0
SBOE district				
1	0	0	43	0
2	0	0	0	0
3	25	0	14	0
4	0	0	14	0
5	38	20	0	20
6	13	0	14	0
7	13	0	0	0
8	0	0	14	0
9	0	20	0	20
10	0	0	0	0
Multiple	0	60	0	60
Years of experience				
< 3	0	0	0	0
3–5	0	40	0	40
6–9	13	0	29	0
≥ 10	63	60	72	60

Note. SBOE = State Board of Education. Column totals may not equal 100% due to rounding or incomplete panelist survey responses.

Appendix G | Subscore Reliability

Table G-1. ELA Subscore Reliability

Grade	Subscore	Reliability
3	CS Cluster	0.59
	KID Cluster	0.73
	LR Cluster	0.55
	Overall Reading	0.85
	Overall Writing	0.68
	Reading Information	0.73
	Reading Literature	0.75
	TTP Cluster	0.61
4	CS Cluster	0.62
	KID Cluster	0.67
	LR Cluster	0.49
	Overall Reading	0.83
	Overall Writing	0.63
	Reading Information	0.71
	Reading Literature	0.71
	TTP Cluster	0.5
5	CS Cluster	0.58
	KID Cluster	0.67
	LR Cluster	0.64
	Overall Reading	0.85
	Overall Writing	0.76
	Reading Information	0.72
	Reading Literature	0.76
	TTP Cluster	0.68
6	CS Cluster	0.73
	KID Cluster	0.73
	Overall Reading	0.88
	Overall Writing	0.58
	Reading Information	0.81
	Reading Literature	0.74
7	CS Cluster	0.74
	KID Cluster	0.7
	Overall Reading	0.88
	Overall Writing	0.6
	Reading Information	0.84

	Reading Literature	0.72
	CS Cluster	0.76
	KID Cluster	0.76
	Overall Reading	0.89
8	Overall Writing	0.6
	Reading Information	0.84
	Reading Literature	0.72
	TTP Cluster	0.52
	CS Cluster	0.76
	KID Cluster	0.66
	Overall Reading	0.87
10	Overall Writing	0.58
	Reading Information	0.82
	Reading Literature	0.7
	TTP Cluster	0.52

Table G-2. Math Subscore Reliability

Grade	Subscore	Reliability
	Measurement & Data	0.79
	Measurement & Data: Area	0.57
3	Number & Operations in Base 10	0.67
	Number & Operations: Fractions	0.63
	Operations & Algebraic Thinking	0.80
	Operations & Algebraic Thinking: Multiplication & Division	0.73
	Number & Operations in Base 10	0.77
	Number & Operations: Using Place Value in Arithmetic	0.71
	Number & Operations: Fractions	0.82
4	Number & Operations: Fractions: Building & Operating on Fractions	0.64
	Number & Operations: Fractions: Decimal Notation	0.72
	Operations & Algebraic Thinking	0.73
	Operations & Algebraic Thinking: Operations with Whole Numbers	0.71
	Measurement & Data	0.72
	Measurement & Data: Volume	0.70
	Number & Operations in Base 10	0.79
5	Number & Operations: Multi-Digit Numbers & Decimals	0.78
	Number & Operations: Fractions	0.79
	Number & Operations: Adding & Subtracting Fractions	0.81
	Number & Operations: Multiplying & Dividing Fractions	0.65
	Expressions & Equations: Proportions & Linear Equations	0.74
6	Expressions & Equations: Equations & Inequalities	0.61
	The Number System	0.77

	The Number System: Common Factors & Multiples	0.68
	The Number System: Rational Numbers	0.63
	Ratios & Proportional Relationships	0.64
	Statistics & Probability	0.53
7	Expressions & Equations	0.52
	Geometry	0.42
	The Number System	0.68
	Ratios & Proportional Relationships	0.48
	Statistics & Probability	0.62
	Statistics & Probability: Probability Models	0.40
8	Expressions & Equations	0.68
	Expressions & Equations: Proportions & Linear Equations	0.56
	Functions	0.70
	Geometry	0.76
	Geometry: Understanding Angles & Angle Measurements	0.78
	Statistics & Probability	0.56
10	Algebra	0.61
	Algebra: Reasoning with Equations & Inequalities	0.47
	Functions	0.52
	Functions: Interpreting Functions	0.51
	Geometry	0.84
	Statistics & Probability	0.56

Table G-3. Science Subscore Reliability

Grade	Subscore	Reliability
5	Earth & Space Science	0.693
	Earth & Space Science: ESS-UEEA	0.584
	Life Science	0.514
	Life Science: LS-DM	0.368
	Physical Science	0.740
	Physical Science: PS-DM	0.615
	Physical Science: PS-EI	0.485
8	Earth & Space Science	0.752
	Earth & Space Science: ESS-DM	0.673
	Life Science	0.737
	Life Science: LS-UEEA	0.627
	Physical Science	0.696
	Physical Science: PS-DM	0.453
	Physical Science: PS-EI	0.453

Appendix H | 2025 KAP Standard Setting – Opening Session Readiness Evaluation Results

Table H-1. Likert-Scale Questions (Percentages)

Question	Bi-grade Panel		Single-Grade Panel	
	Strongly Agree	Agree	Strongly Agree	Agree
The orientation session provided a clear overview of the Bookmark standard-setting process.	48%	48%	30%	50%
I understand my role in the cut score review workshop.	63%	34%	43%	50%
I understand the purpose of the range performance level descriptors (PLDs).	57%	40%	50%	48%
I understand that I will review the range PLDs.	68%	29%	64%	32%
The training on state policy level descriptors was helpful to me.	44%	53%	25%	68%
I understand how to study the items in the Ordered Item Booklet.	35%	58%	14%	43%
I understand that I will receive additional training throughout the workshop.	73%	24%	59%	39%
I understand that the panel will develop descriptions of knowledge and skills differentiating students at different achievement levels.	68%	29%	55%	43%

Table H-2. Yes/No Questions (Counts Only)

Question	Bi-grade Panel		Single-Grade Panel	
	Yes	No	Yes	No
I understand the goals of the workshop.	97	0	44	0
Before I begin work with the range PLDs, I would like additional training on policy level descriptors.	9	88	8	36
I have additional questions on material presented during the opening session that I would like to be answered before I begin the next task.	2	95	0	44

Note. PLD = performance level descriptor.

Appendix I | 2025 KAP Standard Setting – Round 1 Readiness Evaluation Results

Table I-1. Likert-Scale Questions (Percentages)

Question	Bi-grade Panel		Single-Grade Panel	
	Strongly Agree	Agree	Strongly Agree	Agree
I understand how to place the 2025 KAP bookmarks.	61%	36%	48%	48%
I understand that I should set bookmarks independently.	78%	21%	86%	9%
I understand I will have opportunities for further discussion of the bookmarks in Round 2 and 3.	79%	20%	80%	16%
I had an opportunity to ask questions and discuss the meaning of the bookmarks.	75%	24%	80%	16%

Note. KAP = Kansas Assessment Program.

Table I-2. Yes/No Questions (Counts Only)

Question	Bi-grade Panel		Single-Grade Panel	
	Yes	No	Yes	No
I would like additional training on placing my bookmarks for Round 1.	1	95	0	44
I have additional questions that I would like to ask before placing my Round 1 bookmarks.	1	95	0	44