



# KELPA

Kansas English Language Proficiency Assessment

## KELPA Technical Manual

October 2023

Copyright © 2023, Achievement and Assessment Institute, the University of Kansas

## Contents

I.	Statewide System of Standards and Assessments.....	1
I.1	Overview of English Language Standards .....	1
I.2	Test Purposes and Uses .....	1
I.3	Intended Population .....	2
I.4	Overview of Technical Manual Updates .....	2
II.	Assessment System Operations.....	3
II.1	Test Design and Development .....	3
II.2	Content Development.....	3
II.2.1	Rubric Development .....	3
II.2.2	Development of Rater-Training Materials.....	4
II.3	Test Administration and Scoring.....	4
II.3.1	KELPA Teacher Survey.....	5
II.4	Test Security.....	5
II.5	Testing Irregularity .....	5
III.	Technical Quality—Validity .....	7
III.1	Validity Evidence Based on Test Content.....	7
III.2	Validity Evidence Based on Relations to Other Variables .....	9
III.3	Validity Evidence Based on Consequences of Testing .....	12
III.3.1	Teacher Survey.....	12
IV.	Technical Quality—Other.....	15
IV.1	Reliability .....	15
IV.1.1	Test Reliability.....	15
IV.1.2	Classification Consistency and Accuracy .....	17
IV.1.3	Interrater-Agreement Study .....	19
IV.2	Scoring and Scaling .....	24
IV.2.1	Operational Test Results.....	24
IV.2.2	Trend Data .....	37
IV.3	Full Performance Continuum.....	42
V.	Inclusion of All Students.....	44
V.1	Accommodations .....	44
V.1.1	Selection of Accommodations .....	44

V.1.2 Frequency of Accommodations .....	44
V.2 Domain Exemptions.....	45
VI. Academic Achievement Standards and Reporting .....	47
VI.1 Reporting .....	47
VI.1.1 Student Reports.....	47
VI.1.2 Interpretive Guides .....	48
VII. Ongoing Maintenance for KELPA Program .....	49
VII.1 Updates for the 2023 Administration.....	49
VII.2. Plans for Future Administration .....	49
VII.2.1 Multiple Test Forms.....	49
VII.2.2 Improve Reliability .....	49
VII.2.3 Linguistic Process .....	49
References .....	51
Appendix A: 2023 KELPA Teacher Survey .....	52
Introduction .....	52
I. Demographics .....	52
II. Technology.....	53
III. Rater Training for Scoring KELPA Items .....	54
IV. Test-Administration Experience .....	55
V. Student Experience.....	59
VI. Resources.....	60
VII. 2023 KELPA Update .....	61
Appendix B: Summary Results of Teachers’ Responses to Survey Questions .....	62
Appendix C: Responses to Open-Ended Summative Educator Survey Questions .....	67

## Table of Tables

Table III-1. KELPA Blueprint by Percentage of Score Points .....	8
Table III-2. Comparison of Panel Ratings with KELPA Blueprint .....	9
Table III-3. Correlations Between KELPA Domain Scores and KAP English Language Arts (ELA) Scores by Grade.....	10
Table III-4. Correlations Between KELPA Domain Scores and KAP Mathematics Scores by Grade .....	10
Table III-5. Correlations Between KELPA Domain Scores and KAP Science Scores by Grade .....	11
Table III-6. Performance of Proficient English Learners on KAP English Language Arts, Mathematics, and Science Assessments.....	12
Table IV-1. Coefficient Alpha by Domain and Grade or Grade Band .....	16
Table IV-2. Coefficient Alpha for Student Groups by Domain and Grade or Grade Band .....	17
Table IV-3. Classification Consistency (C) and Accuracy (A) by Domain and Grade .....	18
Table IV-4. Available Scoring Methods for Speaking and Writing .....	20
Table IV-5. Numbers of Districts, Schools, and Students Selected for Two Ratings.....	21
Table IV-6. Number of Students with Two Ratings by Domain and Grade or Grade Band .....	21
Table IV-7. Rater Agreement on Writing Items Scored Using the Individual Scoring Method by Grade or Grade Band .....	23
Table IV-8. Rater Agreement on Speaking Items .....	23
Table IV-9. Summary of Quadratic Kappa Classifications .....	24
Table IV-10. 2023 KELPA Participation Rates by State Board of Education (SBOE) District and Grade or Grade Band .....	26
Table IV-11. Percentage of Tested Students by Demographic Characteristic and Grade .....	27
Table IV-12. Scale-Score Descriptive Statistics by Grade for Listening.....	28
Table IV-13. Scale-Score Descriptive Statistics by Grade for Speaking.....	28
Table IV-14. Scale-Score Descriptive Statistics by Grade for Reading .....	29
Table IV-15. Scale-Score Descriptive Statistics by Grade for Writing .....	29
Table IV-16. Demographic Group Scale-Score Descriptive Statistics by Grade for Listening .....	34
Table IV-17. Demographic Group Scale-Score Descriptive Statistics by Grade for Speaking .....	35
Table IV-18. Demographic Group Scale-Score Descriptive Statistics by Grade for Reading.....	36
Table IV-19. Demographic Group Scale-Score Descriptive Statistics by Grade for Writing.....	37
Table IV-20. Number and Percentage of Enrolled and Tested Students by Grade: 2020 Through 2022 ...	38
Table IV-21. Number of Performance Profiles by Grade or Grade Band .....	42
Table IV-22. Domain Performance Profiles by Grade or Grade Band.....	43
Table V-1. Number of Students Using Accommodations by Grade or Grade Band .....	45
Table V-2. Number of Students Exempted for Testing by Domain and Grade .....	46

## Table of Figures

Figure IV-1. 2023 Performance-Level Results for Listening .....	30
Figure IV-2. 2023 Performance-Level Results for Speaking .....	30
Figure IV-3. 2023 Performance-Level Results for Reading .....	31
Figure IV-4. 2023 Performance-Level Results for Writing .....	31
Figure IV-5. Overall Performance-Level Results (2023 Administration) .....	32
Figure IV-6. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Listening .....	39
Figure IV-7. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Speaking .....	39
Figure IV-8. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Reading.....	40
Figure IV-9. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Writing .....	40
Figure IV-10. Comparison of Overall Performance-Level (PL) Results from 2021 Through 2023.....	41

## I. Statewide System of Standards and Assessments

The Kansas English Language Proficiency Assessment (KELPA) is the summative assessment for K–12 English learners (ELs) in Kansas, administered each spring. As part of federal elementary and secondary education legislation for ELs, the test was developed according to the *2018 Kansas Standards for English Learners: Grades K–12* (hereafter referred to as the [2018 Standards](#)). Assessed grades and grade bands include kindergarten, 1, 2–3, 4–5, 6–8, and 9–12. The target student population for KELPA is students identified as ELs in grades K–12.

### I.1 Overview of English Language Standards

The 2018 Standards, developed for grades K–8 and grade bands 9–10 and 11–12, illuminate the critical language, knowledge about language, and language skills that ELs need to be academically successful. The four domains of English language arts (ELA)—listening, speaking, reading, and writing—are the foundation for the 2018 Standards. The 2018 Standards reflect the continual improvement associated with specific, grade-level ELA standards within these four domains. The 2018 Standards are used to support individual students in gaining a level of proficiency in both social English and academic English that allows them to succeed in reaching the grade-level academic standards as quickly as possible. The 2018 Standards also informed the design and content of the new KELPA, first administered in 2020. Refer to the [2020 KELPA Technical Manual](#) (Achievement and Assessment Institute [AAI], 2021a) for more details about the 2018 Standards. The 2023 administration was the fourth of KELPA aligned with the 2018 Standards.

### I.2 Test Purposes and Uses

KELPA is a yearly summative assessment for students in grades K–12 who are identified as not proficient in English, whether they receive English for speakers of other languages (ESOL) services. It is required by the Every Student Succeeds Act (ESSA, 2015), which is the reauthorization of the Elementary and Secondary Education Act (ESEA, 1965). In alignment with the law, KELPA results are used to determine English language proficiency of ELs and assess their progress in acquiring the skills of listening, speaking, reading, and writing in English.

KELPA measures the English language proficiency of ELs to determine who may benefit from receiving the ESOL services and support that ensure students can acquire the language skills to meaningfully participate in educational programs and services. KELPA scores classify ELs' English proficiency into four performance levels (i.e., level 1—beginning, level 2—early intermediate, level 3—intermediate, level 4—early advanced) in each of the four domains and indicate progress toward overall proficiency (i.e., level 1—not proficient, level 2—nearly proficient, level 3—proficient). The proficiency levels determine whether ELs have reached the level of English proficiency that allows them to participate in a standard instructional program in the classroom without additional language support. ELs who demonstrate the English language skills required for engagement with grade-level, academic content instruction at a level comparable to non-ELs (i.e., level 4—early advanced) in all four domains (i.e., listening, speaking, reading, writing) are considered proficient in English language and may exit ESOL program services.

Beyond understanding common English usage, ELs need to understand the language used for grade-level instruction in ELA, mathematics, science, and social studies. The 2018 standards highlight and amplify the critical language, knowledge about it, and skills necessary for ELs to be successful in school.

### I.3 Intended Population

The Kansas State Department of Education (KSDE) is committed to including all eligible ELs in KELPA. Students are identified as ELs when their home or native language is not English and their limitations in the English language may affect their ability to participate in their school's education program. As described, all students in grades K–12 who are identified as ELs must take KELPA, whether or not they receive English language services. For example, parents may waive their student out of ESOL services, but if the student is identified as an EL, he or she is still required to take KELPA. Detailed information about participation in ESOL services and the KELPA program can be found in [ESOL Program Guidance](#) provided by KSDE.

Some ELs may need accommodations for KELPA. When applicable, a student's individualized education program is used to guide accommodations use for KELPA. For more information, refer to the [2020 KELPA Technical Manual](#). A summary of accommodations is provided in Chapter V. Inclusion of All Students in this technical manual.

### I.4 Overview of Technical Manual Updates

A complete technical manual was created for the first year of operational administration in 2020. During the 2020–2021 school year, an independent alignment study was conducted to document validity evidence for KELPA, refer to the [2021 KELPA Technical Manual](#) (AAI, 2021b). This technical manual provides updates for the 2023 administration; therefore, only sections with updated information are included in this manual. For a complete description of KELPA, refer to the [2020 KELPA Technical Manual](#).

## II. Assessment System Operations

This chapter provides updated information about the design, development, administration, and test security of the Kansas English Language Proficiency Assessment (KELPA). For more details (e.g., monitoring test administration), refer to [Chapter II. Assessment System Operations](#) in the *2020 KELPA Technical Manual* (Achievement and Assessment Institute [AAI], 2021a).

### II.1 Test Design and Development

KELPA, part of the Kansas Assessment Program (KAP), is entirely computer based for students in grades 2 through 12. Students in kindergarten and grade 1 take a mostly computer-based exam but also complete a small number of writing items with paper and pencil.

KELPA was designed to be a fixed-form test with one operational form for each domain (i.e., listening, speaking, reading, and writing) and grade level or grade band. All reading and listening items are machine scored, all speaking items are educator scored, and the writing section is composed of both machine- and educator-scored items. The assessments are delivered, in any order of the four domains, through the online test-delivery platform, Kite®.

The University of Kansas's AAI worked with the Kansas State Department of Education (KSDE) to determine the content to be assessed by the KELPA tests for each domain and grade or grade band. The developmental milestones leading to the 2020 KELPA test administration can be found in [Table II-1](#) of the 2020 KELPA Technical Manual, which also provides detailed information about KELPA test blueprints (see [Section II.1.1 Test Blueprints](#)), test design (see [Section II.1.2 Test Design](#)), and test construction (see [Section II.1.3 Test Construction](#)).

### II.2 Content Development

Content development entails various efforts to ensure item quality, including ongoing research into best practices for assessing English learners' proficiency, recruiting highly qualified item writers, developing and providing comprehensive and clear item-writer training materials, conducting item-writer training, and reviewing and revising items. [Section II.2 Content Development](#) in the *2020 KELPA Technical Manual* includes detailed descriptions of the typical procedures for various stages of content development:

- [Section II.2.1 Passage Development](#)
- [Section II.2.2 Item Writing](#)
- [Section II.2.3 Item Review](#)

This section provides updated information about the development of both the rubric and the rater-training materials.

#### II.2.1 Rubric Development

KELPA rubrics developed for the 2020 administration were used in 2021. Refer to [Section II.2.4 Rubric Development](#) in the 2020 KELPA Technical Manual for detailed activities of rubric development by phase. To support rater use of the rubrics in kindergarten and grade 1, supplementary documents were added to the rater-training materials to provide additional, more specific guidance on using the writing



rubrics in those grades. These supplemental documents were also developed in 2020 and used in the 2020 through 2023 administrations.

## II.2.2 Development of Rater-Training Materials

This section describes the development of updated rater-training materials for the 2023 KELPA administration, the final year of the staged roll-out to ensure all constructed-response (CR) items on the assessment include item-specific rater-training materials.

In 2021, the rater-training materials included one set of materials (sets include anchor, calibration, and practice responses) for one writing and one speaking prompt per grade or grade band. For the 2022 administration, these materials were expanded to include additional prompts. For the 2023 administration, rater-training materials were further expanded to include all writing and speaking prompts administered on the assessment.

The development process for 2023 materials was the same as what was used for the 2022 materials, see the [2022 KELPA Technical Manual](#) for details. AAI content-development staff selected responses for all sets and wrote explanations for the anchor-set responses. During external reviews, educators reviewed anchor and calibration sets, and KSDE staff reviewed all sets. AAI content-development staff used synchronous and asynchronous feedback to select and determine any needed replacements, which were reviewed and approved by KSDE. The materials were posted on Educator Portal prior to the 2023 KELPA administration window.

## II.3 Test Administration and Scoring

The 2023 KELPA testing window was open to students from January 30 through March 10, 2023. Educators were able to enter scores for CR items until March 31, 2023. Additional information about scoring can be found in the [KELPA Scoring Manual](#). For an overview of KELPA administration and scoring, refer to the introductory paragraphs of [Section II.3 Test Administration and Scoring](#) in the *2020 KELPA Technical Manual*.

Kansas uses a train-the-trainer model in which District Test Coordinators (DTCs) receive training directly from KSDE and, in turn, train educators in their local school districts in test administration and scoring. DTCs are responsible for training educators in scoring CR items in speaking and writing, as well as training test-administration staff on test security and ethics. For more information about this model and training details, refer to [Section II.3.1 Test-Administrator and Scorer Training](#) of the *2020 KELPA Technical Manual*. The provided training webinars, recorded and posted on site, are updated every year. The training slides, frequently asked questions, and responses to these questions are also posted on the [DTC Virtual Training](#) site.

The standardized test-administration procedures provided for districts, schools, and teachers are described in the [2022–2023 KELPA Examiner’s Manual](#) (*Examiner’s Manual* hereafter). The *Examiner’s Manual* also provides guidance and procedures related to the administration of KELPA in 2022–2023. For example, it includes procedures and information needed to prepare students and administrators before, during, and after KELPA (sections 4, 5, and 6, respectively). A summary of these details is provided in [Section II.3.2 Test-Administration Procedures](#) of the *2020 KELPA Technical Manual*.

### II.3.1 KELPA Teacher Survey

At the beginning of the KELPA testing window, KSDE sent out a notification about the KELPA teacher survey through KSDE email distribution lists to encourage educators to participate. At the same time, an announcement about the teacher survey was posted in the Educator Portal. The purpose of the survey was to collect information about educators' experience with KELPA. The survey was available in the Educator Portal through March 31, 2023. The survey (see Appendix A) included questions about educators' backgrounds and their experience with Kite, scoring, test administration, student experience, supporting materials (e.g., the [2022–2023 KELPA Examiner's Manual](#), [KELPA Test Administration and Scoring Directions](#) for speaking and writing, etc.), learning and instruction in 2022–2023, and the utility of KELPA. A total of 101 educators responded to the survey. Results of the teacher survey are included in III.3.1 Teacher Survey [of](#) the current manual, as well as Appendix B and Appendix C.

## II.4 Test Security

Test security is maintained by protecting the integrity and confidentiality of test materials, test-related data, and personally identifiable information. For a summary of KSDE's plan for ensuring the security and confidentiality of state testing materials, refer to [Section II.5 Test Security](#) of the *2020 KELPA Technical Manual*. For more details about security requirements, refer to the [Kansas Assessment Fact Sheet: Test Security and Ethics](#) and the [Kansas State Department of Education Test Security Guidelines](#). [Sections II.5.1](#) through [II.5.4](#) of the *2020 KELPA Technical Manual* provide detailed information about and requirements for test-materials security, test-related data security, security of personally identifiable information, and accommodations-related security.

## II.5 Testing Irregularity

During the Spring 2023 KELPA test-administration window, KSDE received a total of 58 test resets. A test reset delivers the same test (domain) and wipes the previous responses. This requires approval from the state. Test administrators and coordinators are trained to report test-administration irregularities. During the operational window, monitoring of testing data was conducted by AAI, which oversees and manages the Kite system. ATS conducted data validation daily to monitor system usage and identify testing irregularities. A dashboard of testing activities is available for Educator Portal for administrators in the field to monitor and record activities for the KAP program. The dashboard records all system usage, including 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 (which activate a test if needed, without wiping previous responses and does not need KSDE approval).

### **Examples of testing irregularities include the following:**

- Fast test-taking behavior (i.e., students finished a test section in a short amount of time)
  - Typically only requires a reactivation and not a reset for KELPA
- Irregular testing time (i.e., a test session started or ended outside of school hours),
  - Recorded by the system and can be viewed in the dashboard in Kite Educator Portal
- Tests reactivated by users (i.e., test administrators) due to student enrollment or demographic data errors
  - Recorded by the system and can be viewed in the dashboard
- Student clicks through a test and then submits without answering any questions.

- Requires a test reactivation, which can be done by the testing coordinator
- The student was caught cheating.
  - State approval is required for a test reset. The state may require the user to enter a special circumstance code (SC-28). The test proctor will notify the DTC, who will call KSDE for guidance.
  - If a test reset is **not** approved, the DTC will enter the SC code. This makes the test invalid; the student will be recorded as not tested, and it will result in lower participation for the specific school/district.
- The student took someone else's test by accident.
  - Requires a test reset or moving the data to the correct test. Both options require approval from KSDE.
- The educator scored the wrong student and needs the scores moved to the correct student.
  - Requires approval from KSDE to move the scores
- The student's personal need profile (PNP) was not set up correctly before testing.
  - Requires approval from KSDE to reset the test

### III. Technical Quality—Validity

*Standards for Educational and Psychological Testing* defines *validity* as “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). There are five sources of evidence to consider when evaluating test-score validity (APA et al., 2014): (a) test content, (b) response processes, (c) internal test structure, (d) relationships between test scores and other variables, and (e) consequences of testing. The Kansas English Language Proficiency Assessment (KELPA) test forms in 2023 were the same as the operational forms from 2020 and 2022; therefore, the evidence from the model calibration and differential item functioning analysis did not need to be updated. For details about validity evidence based on internal structure and other additional evidence, refer to [Chapter III. Technical Quality—Validity](#) in the *2020 KELPA Technical Manual* (Achievement and Assessment Institute [AAI], 2021a). This chapter presents validity evidence collected or evaluated during the 2022–2023 school year.

#### III.1 Validity Evidence Based on Test Content

Validity evidence based on test content is used to demonstrate that the content of the test is related to the specific content domains the test was intended to measure. The interpretation and use of KELPA results rely on the correspondence between items and the [2018 Standards](#), as well as between the test and test blueprint. The external, independent KELPA alignment study was conducted by the Human Resources Research Organization (HumRRO) with participation of Kansas educators in spring 2021 to examine the extent of alignment among KELPA, the 2018 Standards, and the academic content standards (Sinclair et al., 2021).

Two of the recommendations provided by the independent alignment study are to (1) present the test blueprint in terms of percentage of score points and (2) specify how constructed-response items contribute to the “language in speaking and listening” cluster in the speaking domain. AAI reviewed the original test blueprint (presented in [Chapter II. Assessment System Operations](#) in the *2020 KELPA Technical Manual*) and took into account dual alignment for constructed-response items. Table III-1 shows the KELPA blueprint in the updated format.

Table III-1. KELPA Blueprint by Percentage of Score Points

Domain	Cluster	Percentage of score points with range of +/- 10%					
		GK	G1	G2–3	G4–5	G6–8	G9–12
Reading	Reading Foundations	70	60	25	20	10	10
	Language in Reading	20	30	20	35	35	40
	Discourse Comprehension	5	5	45	45	45	40
	Craft & Structure	5	5	10	0	10	10
Listening	Comprehension & Collaboration	60	80	75	70	80	70
	Presentation of Knowledge & Ideas	5	10	15	15	10	15
	Language in Speaking & Listening	35	10	10	15	10	15
Speaking	Comprehension & Collaboration	30	30	30	30	60	45
	Presentation of Knowledge & Ideas	60	50	50	45	30	45
	Language in Speaking & Listening	10	20	20	30	10	10
Writing	Language in Writing	100	100	70	65	60	60
	Production of Writing	0	0	30	35	40	40

After restructuring the test blueprint, AAI content-development staff conducted a post-alignment reconciliation. AAI content-development staff reviewed and realigned items marked as partially aligned from the independent alignment study panel and assigned final alignment to the items. AAI content-development staff also added secondary alignment for constructed-response items when needed to reflect skills described in the rubric.

The AAI psychometric team analyzed the independent panel ratings (with updates from the aforementioned reconciliation) and evaluated blueprint coverage. All grades and domains aligned to the test blueprint, except for grade K and grade 1 reading. Table III-2 shows the panel ratings compared to the blueprint. It shows that, based on panel ratings, the KELPA forms include a higher percentage of reading-foundations items than expected from the test blueprint. When comparing original metadata and panel ratings, it reveals that panelists considered R1 and R11 items (in metadata) to RF3 for grade 1 and R11 and R12 items (in metadata) to RF3 for grade K. Reading Foundations Standard 3 (RF3) deals with knowing and applying grade-level phonics and word-analysis skills when decoding words. It is a standard that deals with foundational, word-level language decoding and understanding the basic building blocks of word parts.

Standard R1 deals with meaning making when the answer is explicitly stated in the text. It is less about word-level decoding and more about understanding the meanings of words in the context of phrases, sentences, and whole paragraphs.

Standard R11 deals with applying context-cueing systems to understand words as they are used in context. Standard R12 deals with parsing multiple-meaning words, connotations, denotations, nuance, and shades of meaning. This misalignment may have been caused by the focus on the science of reading and the high value placed on foundational reading skills such as decoding and word analysis. However, when these alignments were analyzed by content specialists and reading experts, the items were clearly

not well-aligned to the foundational reading standard RF3 and much more closely aligned to R11 and R12, because the items asked about how the words operate in the context of a whole phrase, sentence, or paragraph. Rather than phonics and decoding, the linguistic focus of these items was semantics, or meaning making.

*Table III-2. Comparison of Panel Ratings with KELPA Blueprint*

Cluster		Grade K		Grade 1	
		Blueprint (%)	Panel ratings (%)	Blueprint (%)	Panel ratings (%)
Reading	Reading Foundations	70	89	60	72
	Language in Reading	20	11	30	12
	Discourse Comprehension	5	0	5	12
	Craft & Structure	5	0	5	4

Based on the result of the re-analyses for panel ratings, AAI staff will work with KSDE to evaluate whether the KELPA blueprint for reading grade K and grade 1 should be adjusted in a future administration when KELPA forms are updated.

## III.2 Validity Evidence Based on Relations to Other Variables

The external validity evidence is defined as “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 three types of external evidence are convergent, discriminant, and criterion related (either predictive or concurrent).

*Convergent evidence* is provided by relationships among students’ performance on different assessments measuring similar constructs. *Discriminant evidence* is provided by relationships among students’ performance on different assessments measuring different constructs. *Criterion-related evidence* is provided by relationships between students’ test scores on one test and those on another test of a related attribute (Cronbach, 1951; Messick, 1989).

The external assessments used in this study are the Kansas Assessment Program (KAP) English language arts (ELA) and mathematics assessments, which are administered annually to students in grades 3–8 and 10, as well as the KAP science assessment, which is administered annually to students in grades 5, 8, and 11. The Pearson product-moment correlations between KELPA-domain scale scores and KAP ELA, mathematics, or science scale scores can provide validity evidence based on relations to other variables. The effect size is considered small if a correlation coefficient is less than .30, large if equal to or greater than .50, and medium if in between (Cohen, 1988). Relationships between KAP-subject scale scores and KELPA-domain scale scores were examined because English learners’ (ELs’) proficiency in each KELPA domain may have a different impact on their performance in the grade-level academic tests.

Table III-3 presents correlation coefficients between KELPA domain scores and KAP ELA scores. The strongest correlations were between KAP ELA and the KELPA reading domain, ranging from .51 (grade 8) to .63 (grades 4 and 6); the weakest correlations were observed between ELA and the speaking domain, ranging from .21 (grade 7 and grade 8) to .31 (grade 3). Correlation coefficients between the KAP ELA and KELPA speaking domain across grades were small (except in grades 3 and 4). For relationships

between KAP ELA and KELPA listening, reading, and writing, medium to large correlation coefficients were found across grades.

*Table III-3. Correlations Between KELPA Domain Scores and KAP English Language Arts (ELA) Scores by Grade*

Grade	Correlation between KAP ELA and domain			
	Listening	Speaking	Reading	Writing
3	.48	.31	.62	.57
4	.54	.30	.63	.56
5	.48	.24	.57	.46
6	.54	.29	.63	.52
7	.46	.21	.57	.43
8	.45	.21	.51	.38
10	.40	.22	.53	.39

Table III-4 presents correlations between KELPA domain scores and KAP mathematics scores. Compared to the relationships with KAP ELA, relationships between KELPA domain scores and KAP mathematics scores were weaker in all domains. The strongest correlation was between KAP mathematics and KELPA reading domain, ranging from .25 (grade 10) to .54 (grade 3); the weakest correlation was between KAP mathematics and KELPA speaking domain, ranging from .07 (grade 10) to .31 (grade 3). Relationships between KAP mathematics and KELPA were weakest for grade 10.

*Table III-4. Correlations Between KELPA Domain Scores and KAP Mathematics Scores by Grade*

Grade	Correlation between KAP mathematics and domain			
	Listening	Speaking	Reading	Writing
3	.46	.31	.54	.55
4	.44	.23	.47	.46
5	.37	.15	.41	.40
6	.41	.25	.46	.39
7	.34	.18	.40	.30
8	.34	.21	.39	.31
10	.20	.07	.25	.19

Table III-5 presents correlations between KELPA domain scores and KAP science scores. The strongest correlation was between KAP science and reading scores, ranging from .34 (grade 11) to .50 (grade 5); the weakest correlation was between science and speaking scores, ranging from .09 (grade 11) to .23 (grade 5). Correlations between KAP science and KELPA scores were weakest for grade 11.

*Table III-5. Correlations Between KELPA Domain Scores and KAP Science Scores by Grade*

Grade	Correlation between KAP science and domain			
	Listening	Speaking	Reading	Writing
5	.47	.23	.50	.40
8	.34	.17	.43	.23
11	.25	.09	.34	.22

Table III-6 presents student performance on KAP ELA, mathematics, and science for proficient KELPA students. More proficient ELs in lower grades scored proficient in KAP ELA and mathematics compared to students at higher grades. For example, 32% of proficient ELs in grade 3 scored at level 3 or level 4 (proficient) in KAP ELA, compared to 12% of proficient ELs in grade 7. Proficient ELs at grade 10 had the lowest performance in KAP ELA and mathematics: only 5% of proficient ELs scored at level 3 or level 4 on KAP ELA, and only 3% of proficient ELs scored at level 3 or level 4 on mathematics. For science, proficient ELs at grade 8 had the lowest performance: only 7% of proficient ELs scored at level 3 or level 4 on KAP science.



Table III-6. Performance of Proficient English Learners on KAP English Language Arts, Mathematics, and Science Assessments

Grade	Proficient English learners (ELs)											
	KAP English language arts				KAP mathematics				KAP science			
	Proficient ELs (n)	Level 3 or 4 (proficient) (%)	Level 2 (%)	Level 1 (%)	Proficient ELs (n)	Level 3 or 4 (proficient) (%)	Level 2 (%)	Level 1 (%)	Proficient ELs (n)	Level 3 or 4 (proficient) (%)	Level 2 (%)	Level 1 (%)
3	504	32	53	16	509	55	34	11	—	—	—	—
4	631	32	60	7	636	29	58	14	—	—	—	—
5	543	19	45	36	546	14	47	39	543	26	43	31
6	263	22	42	35	267	15	45	40	—	—	—	—
7	242	12	39	49	248	11	61	28	—	—	—	—
8	252	8	44	48	256	9	25	66	255	7	27	66
10	227	5	42	53	230	3	26	71	—	—	—	—
11	—	—	—	—	—	—	—	—	112	12	22	66

### III.3 Validity Evidence Based on Consequences of Testing

Details about validity evidence based on consequences of testing are described in [Section III.5 Validity Evidence Based on Consequences of Testing](#) in the *2020 KELPA Technical Manual*. A teacher survey collected an additional piece of evidence based on consequences of testing during the 2023 KELPA administration. Appendix B and Appendix C present results of the teacher survey for selected-response and open-ended-response questions. Responses to some survey questions indicate that most participating educators believed that the content of KELPA measured important English language proficiency knowledge, skills, and abilities (91%,  $n = 98^1$ ).

#### III.3.1 Teacher Survey

In the current document, the results in Table B-1 show that half (50%) of the participating educators who responded to the survey were teachers (i.e., classroom, Title 1, special education, ELs). Many of these educators had 10 or more years of experience in ELA (69%), in mathematics (59%), in science (48%), and/or with ELs (66%; see Table B-3). They were well distributed across different grades or grade bands (see Table B-2), ranging from 7% (grades 9–12) to 21% (grades 2–3).

<sup>1</sup> 2% of participants did not respond to this survey question.

The results in Table B-4 show the device type used by most assessed students. Chromebook and iPad were dominantly used for kindergarten (32% and 35% respectively) and grade 1 (33% for both device types). For older grade levels, Chromebook gradually got adopted more than iPad. For example, for kindergarten, the Chromebook and iPad percentages were 32% and 35%, whereas for grades 9–12, the corresponding percentages were 31% and 2%, respectively. It was noticeable that, as grade levels increased, the percentages of participants that did not respond to this question increased, ranging from 26% (grades 2–3 and grades 4–5) to 61% (grades 9–12).

Most educators somewhat agreed or agreed that the rater-training materials (see Table B-5) helped to apply rubrics for scoring students' responses to both speaking (92%) and writing items (91%) and that the length of the state scoring window was sufficient (95%). As for the rater-training workshops (see Table B-6), most educators somewhat agreed or agreed that the local rater training helped to understand the scoring rubrics (72%), and to know how to use the scoring rubrics (72%). They somewhat agreed or agreed that the local rater training provided useful information for their roles as raters (73%). Moreover, they somewhat agreed or agreed that the training was well organized (75%) and appropriately timely (77%). Most educators somewhat agreed or agreed that the KSDE-published rater-training materials were easy to use (84%) and helped to score responses confidently (87%). Table B-7 shows that the vast majority of educators responded positively about their test-administration experience for all four domains, regarding the appropriateness of domain test length (91–93%), the clarity of test instruction (96–98%), and the helpfulness of test instruction to students (93–95%).

Table B-8 shows that the educators highly rated their test-administration experience in general. Specifically, they somewhat agreed or agreed that they are confident in their ability to administer KELPA (100%), that the relevant training prepared them for the responsibilities of a test administrator (97%), and that the District Test Coordinator or Building Test Coordinator training sessions provided across the state were helpful (77%<sup>2</sup>).

Table B-9 shows the educators' responses about student experience. Particularly, most educators somewhat agreed or agreed that the content of KELPA measured important English language proficiency knowledge, skills, and abilities (91%); that their students had access to all necessary accessibility supports to participate in the assessment (94%); that, in general, ELs classified as proficient according to their KELPA scores can fully access grade-level academic content (91%); and that ELs classified as not proficient are not able to do so without ESOL services (81%).

Table B-10 shows that the educators responded positively to questions about the utility of the [2022–2023 KELPA Examiner's Manual](#) (98%), the [KELPA Scoring Manual](#) (97%), the [Kite® Practice Test Guide for Educators](#) (87%), and the *KELPA Test Administration and Scoring Directions*<sup>3</sup> for both speaking and writing documents (98%).

This year, the educators were also surveyed about their experience with the update to the summative KELPA program. As shown in Table B-11,<sup>4</sup> the educators somewhat agreed or agreed that the KELPA school and district reports were overall helpful (67%) and easy to understand (74%), and that the school-specific results on the KELPA district report were helpful (66%).

The teacher survey included some open-ended questions to collect educators' feedback on a variety of topics. These topics included the use of Kite Suite applications with KELPA summative assessments in

---

<sup>2</sup> 22% of the participants responded that this question was not applicable.

<sup>3</sup> Secure KELPA scoring manual is available in the Educator Portal.

<sup>4</sup> More than 21% of the participants responded that the questions in this section were not applicable.

2022–2023 administration, suggestions to help improve educators’ ability to administer KELPA, and the use of the new KELPA school and district reports.

Refer to Appendix C: Responses to Open-Ended Summative Educator Survey Questions for more information about educators’ responses.

## IV. Technical Quality—Other

This chapter provides updated evidence related to the technical quality of the Kansas English Language Proficiency Assessment (KELPA) administered in 2023, including reliability-related evidence, a summary of test results, and a description of ongoing program improvement. For technical-quality-related evidence, refer to [Section IV.2 Fairness and Accessibility](#) and [Section IV.4 Full Performance Continuum](#) in the *2020 KELPA Technical Manual* (Achievement and Assessment Institute [AAI], 2021a). Quality-control steps were elaborated in [Section IV.3.5 Quality-Control Checks](#) in the *2020 KELPA Technical Manual*.

### IV.1 Reliability

*Reliability* refers to the degree of consistency of students' test scores across repeated measures. When a test is reliable, a student's test scores from multiple standard administrations under the same testing conditions are relatively stable. Reliability is typically estimated from student-response data rather than calculated directly because it is not possible for a student to take the same test multiple times without any changes to the testing conditions. According to the *Standards for Educational and Psychological Testing* (American Psychological Association et al., 2014):

The term *reliability* has been used in two ways in measurement literature. First, the term has been used to refer to the reliability coefficients of classical test theory, defined as the correlation between scores on two equivalent forms of the test, presuming that taking one form has no effect on performance on the second form. Second, the term has been used in a more general sense, to refer to the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported (e.g., in terms of standard errors, reliability coefficients per se, generalizability coefficients, error/tolerance ratios, item response theory (IRT) information functions, or various indices of classification consistency). (p. 33)

The reliability estimates for KELPA are reported in two ways: reliability coefficients from classical test theory (CTT) and IRT information functions combined with conditional standard error of measurement. CTT reliability coefficients are sample dependent and were updated using the 2023 data. IRT reliability does not change by test sample, only by test form. Because the same test forms were used from 2020 to 2023, IRT reliability is not provided in this section. For detailed information about IRT reliability, refer to [Section IV.1 Reliability](#) of the *2020 KELPA Technical Manual*. For CTT reliability coefficients, the student-group reliabilities were also calculated. Indices of classification consistency and accuracy of different domain performance levels, as well as interrater agreement on speaking and writing constructed-response (CR) items, are also provided in this section of the current manual. For information about the fairness and accessibility of KELPA, refer to [Section IV.2 Fairness and Accessibility](#) of the *2020 KELPA Technical Manual*.

#### IV.1.1 Test Reliability

Because KELPA uses only one fixed form for each domain test at each grade or within each grade band, the coefficient alpha index of internal consistency (Cronbach, 1951) from CTT is calculated. The formula (i.e., Equation IV-1) for the coefficient alpha index is:

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_x^2} \right], \quad (\text{IV-1})$$

where  $k$  is the number of items on the test form,  $\sigma_i^2$  is the variance of item  $i$ , and  $\sigma_x^2$  is the total test variance. KELPA reliability coefficients by domain and grade or grade band can be found in Table IV-1. Reliabilities of the KELPA domain tests were adequate, with indices ranging from .80 to .97 across most grades or bands and domains. The exceptions were in kindergarten for reading (.69) and writing (.74). Test length and test reliability are closely related, and shorter tests are usually less reliable. Compared to other domains, kindergarten reading and writing tests had lower reliabilities because these tests had the fewest score points among all grade levels. [Table II-13](#) in the *2020 KELPA Technical Manual* indicates the test lengths and total score points for all domain tests.

*Table IV-1. Coefficient Alpha by Domain and Grade or Grade Band*

Grade or grade band	Listening $\alpha$	Speaking $\alpha$	Reading $\alpha$	Writing $\alpha$
K	.86	.93	.69	.74
1	.86	.93	.89	.81
2–3	.89	.93	.90	.86
4–5	.89	.94	.84	.84
6–8	.87	.95	.84	.87
9–12	.90	.97	.85	.80

#### *IV.1.1.1 Student-Group Reliability*

Reliability estimates were also calculated by the student group and are presented in Table IV-2. Results show that the student-group reliabilities were similar within a domain and at most grades or grade bands; the exceptions were kindergarten in reading and writing, where reliability coefficients for student groups were lower (consistent with the domain-level coefficient alphas). Also, the student-group reliabilities were similar to the overall reliabilities, with most estimates in the .80s to .90s; reading in kindergarten (mostly in the .70 range or lower) and writing in kindergarten (mostly in the .70 range) and in grade band 9–12 (group with disabilities only,  $\alpha = .73$ ) had lower reliabilities. The sample size of each student group can be found in Section IV.2.1.1 Test-Enrollment Data of the current document.

Table IV-2. Coefficient Alpha for Student Groups by Domain and Grade or Grade Band

Domain and grade or grade band	Coefficient $\alpha$						
	Female	Male	White	Non- White	Hispanic	Non- Hispanic	SWD SWOD
Listening							
K	.85	.86	.85	.87	.85	.89	.89
1	.84	.86	.86	.85	.86	.86	.86
2–3	.88	.89	.88	.90	.88	.90	.89
4–5	.88	.90	.89	.90	.89	.91	.88
6–8	.88	.87	.87	.89	.87	.89	.84
9–12	.90	.90	.90	.90	.90	.90	.86
Speaking							
K	.93	.93	.93	.94	.93	.94	.95
1	.93	.93	.93	.94	.93	.94	.94
2–3	.94	.93	.93	.93	.93	.94	.93
4–5	.95	.94	.94	.94	.94	.95	.92
6–8	.96	.95	.95	.96	.95	.95	.95
9–12	.97	.97	.97	.97	.97	.97	.97
Reading							
K	.68	.70	.65	.76	.62	.79	.68
1	.88	.89	.88	.90	.88	.90	.86
2–3	.90	.90	.90	.91	.90	.91	.88
4–5	.84	.85	.84	.85	.84	.86	.81
6–8	.83	.85	.84	.86	.84	.85	.81
9–12	.84	.86	.85	.85	.85	.87	.82
Writing							
K	.72	.74	.72	.76	.71	.78	.77
1	.79	.82	.80	.82	.80	.82	.81
2–3	.86	.87	.86	.87	.86	.88	.85
4–5	.84	.84	.84	.85	.83	.86	.82
6–8	.87	.87	.87	.88	.87	.89	.84
9–12	.81	.79	.80	.81	.80	.82	.73

Note. SWD = students with disability; SWOD = students without disability.

#### IV.1.2 Classification Consistency and Accuracy

When an assessment primarily uses achievement or proficiency levels to report test results, accuracy and consistency of classification into different proficiency levels become key indicators of the quality of the assessment. As described by 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). The coefficients for classification

consistency and accuracy range from 0 to 1, with 0 representing classifications that are not consistent or accurate and 1 representing perfectly consistent or accurate classifications.

Detailed descriptions of the calculation of two indices can be found in [Section IV.1.3 Classification Consistency and Accuracy](#) in the *2020 KELPA Technical Manual*. The results for classification consistency and accuracy for three cuts are presented in Table IV-3. The classification consistency and accuracy of the level-4 cut are particularly important for proficiency classification because students must be at level 4 in all four domains to be considered proficient overall. Classification-consistency indices for the KELPA domain tests ranged from .69 to .98 across most cuts and grades or grand bands. Classification-accuracy indices for the KELPA domain tests ranged from .76 to .99 across most cuts and grade levels or bands.

*Table IV-3. Classification Consistency (C) and Accuracy (A) by Domain and Grade*

Domain and grade	Cut-score category					
	1 vs. 2, 3, 4		1, 2 vs. 3, 4		1, 2, 3 vs. 4	
	C	A	C	A	C	A
<b>Listening</b>						
K	.93	.95	.91	.94	.76	.82
1	.94	.96	.88	.92	.81	.86
2	.97	.98	.92	.94	.87	.91
3	.98	.98	.95	.96	.90	.93
4	.96	.97	.95	.96	.85	.90
5	.96	.97	.95	.96	.85	.90
6	.95	.97	.94	.96	.82	.87
7	.95	.96	.93	.95	.84	.89
8	.95	.96	.94	.96	.81	.87
9	.92	.95	.92	.94	.88	.91
10	.92	.95	.91	.94	.89	.92
11	.93	.95	.93	.95	.84	.89
12	.93	.95	.93	.95	.85	.90
<b>Speaking</b>						
K	.93	.95	.90	.93	.81	.85
1	.96	.97	.93	.95	.77	.84
2	.97	.98	.95	.96	.81	.87
3	.97	.98	.96	.97	.82	.88
4	.98	.99	.96	.97	.87	.91
5	.97	.98	.96	.97	.78	.85
6	.97	.98	.95	.97	.83	.88
7	.97	.98	.95	.97	.83	.88
8	.97	.98	.96	.97	.75	.83
9	.96	.97	.96	.97	.91	.94
10	.96	.97	.95	.97	.92	.94
11	.96	.97	.96	.97	.90	.93
12	.96	.97	.96	.97	.87	.91

Domain and grade	Cut-score category					
	1 vs. 2, 3, 4		1, 2 vs. 3, 4		1, 2, 3 vs. 4	
	C	A	C	A	C	A
Reading						
K	.69	.77	.83	.89	.92	.95
1	.86	.90	.87	.91	.92	.94
2	.87	.90	.89	.92	.89	.93
3	.90	.93	.90	.93	.87	.91
4	.91	.93	.85	.89	.83	.88
5	.90	.93	.84	.89	.81	.86
6	.93	.95	.85	.89	.83	.88
7	.91	.93	.85	.89	.82	.87
8	.91	.94	.85	.90	.78	.83
9	.86	.90	.87	.91	.90	.93
10	.86	.90	.87	.90	.88	.91
11	.86	.90	.86	.90	.86	.90
12	.87	.91	.85	.90	.85	.89
Writing						
K	.84	.89	.77	.84	.88	.91
1	.94	.96	.84	.89	.75	.79
2	.92	.94	.87	.91	.79	.85
3	.93	.95	.88	.92	.75	.79
4	.93	.95	.90	.93	.77	.84
5	.95	.96	.88	.92	.73	.80
6	.95	.97	.90	.93	.74	.80
7	.95	.97	.87	.91	.74	.79
8	.96	.97	.86	.90	.70	.76
9	.87	.91	.79	.85	.78	.84
10	.87	.91	.80	.86	.79	.85
11	.85	.90	.78	.85	.74	.79
12	.86	.90	.77	.84	.71	.76

*Note.* Categories 1, 2, 3, and 4 represent proficiency levels 1, 2, 3, and 4, respectively.

#### IV.1.3 Interrater-Agreement Study

The interrater-agreement study provides reliability and validity evidence for the educator-scored test items. KELPA CR item scores ranged from 0 to 3 for both speaking and writing. Refer to [Table II-13](#) in the *2020 KELPA Technical Manual* for the number of educator-scored items for speaking and writing by grade or grade band. Within the same grade or grade band in each domain of speaking and writing, holistic rubrics were used to rate CR item responses instead of item-specific rubrics. The rater training provided at local schools and districts, as well as the training materials provided by KSDE, supplied educators with the knowledge and skills needed to apply the rubrics. The scoring accuracy of CR items, which are scored by educators, relies on consistent and appropriate application of the scoring rubrics. Therefore, it is worthwhile to evaluate whether teachers applied the rubrics consistently—the



interrater-agreement study results can help identify further improvements to training materials—and to examine how much raters agreed or disagreed with each other on their ratings for each of the CR items.

*IV.1.3.1 Data-Collection Method*

An interrater-agreement study of KELPA writing and speaking CR items was conducted during the 2023 KELPA scoring window (January 29–March 29, 2023). Two options were provided to collect second ratings: Kite® Educator Portal scoring interface or a spreadsheet for targeted school districts. The Kite Educator Portal scoring interface was used for individual raters to manually score questions that are not machine scored, and the spreadsheet option was used for school districts to enter information for a roster of students in batches. To allow two scorers to enter scores for the same student response, students selected for second ratings had two scoring tabs in Educator Portal for all CR items. Score of record (used in score reporting, i.e., the first score entered; refer to [Section IV.3.1.2 Educator Scoring](#) of the *2020 KELPA Technical Manual* for more information about how scores were entered) for operational scoring remained the same for all students regardless of whether a student was selected for second rating. District Test Coordinators (DTCs) were responsible for monitoring the process for collecting second ratings from selected educators in their district. Table IV-4 shows available scoring methods for both first and second raters in speaking and writing. Note that for speaking, in addition to individual or paired/group scoring, educators could also choose deferred scoring (by listening to audio playback) or simultaneous scoring (by sitting next to students during testing).

*Table IV-4. Available Scoring Methods for Speaking and Writing*

Writing	Speaking	
	Option 1	Option 2
Individual scoring or paired/group scoring	Individual scoring or paired/group scoring	Deferred scoring or simultaneous scoring

In addition to the second scores, information collected through the user interface of Educator Portal also included:

- **Scoring method, first rating:** Users may select individual scoring (i.e., each scorer works independently) or paired/group scoring (i.e., scorers work in pairs or a small group).
- **Speaking scoring options, first rating:** Users may select simultaneous scoring (i.e., scoring items in the moment that students are responding) or deferred scoring (i.e., scoring items later by listening to the recordings).
- **Designated scorer, first rating:** Defaults to the user who is logged in; users may change the name of the scorer, if scored by another user.
- **Scoring method, second rating:** Users may select individual or paired/group scoring.
- **Speaking scoring options, second rating:** Users may select simultaneous or deferred scoring.
- **Designated scorer, second rating:** Defaults to the user who is logged in; users may change the name of the scorer, if scored by another user.

*IV.1.3.2 Sampling*

A sample of students taking KELPA for the 2023 administration was selected to receive second ratings for their speaking and writing CR items. Samples selected for two ratings were identified at the

beginning of the testing window when all school districts completed KELPA test registration. Selected students received two ratings for each CR item, with a target sample size of approximately 500 students per grade. A random sample of 15% of registered kindergarten and grade 1 students was selected. A random sample of 14% of registered students in grades 2–12 was also selected. Table IV-5 shows the numbers of districts, schools, and students selected for the two ratings.

*Table IV-5. Numbers of Districts, Schools, and Students Selected for Two Ratings*

Grade or grade band	No. of districts	No. of schools	No. of students
K	31	137	485
1	35	130	522
2–3	38	173	724
4–5	35	146	592
6–8	41	100	749
9–12	42	64	824

Data obtained at the end of the window for hand scoring speaking and writing items were used for rater-agreement analyses. Only an exceedingly small percentage (0–2%) of responses with two ratings were collected using the paired/group scoring method for both writing and speaking. For speaking responses scored individually, 0–3% of these responses were simultaneously scored. Sample sizes, both for paired/group scoring in writing and speaking and simultaneous scoring for speaking, were not sufficient to make meaningful statistical inferences. Therefore, Table IV-6 shows the number of student responses per item using the individual scoring method for writing, as well as the number of student responses per item using the combination of individual and deferred scoring methods for speaking.

*Table IV-6. Number of Students with Two Ratings by Domain and Grade or Grade Band*

Grade or grade band	Number of student responses per item	
	Writing: Individual scoring	Speaking: Combination of individual and deferred scoring
K	438–440	349–358
1	475–479	387–393
2–3	649–652	563–572
4–5	530–537	465–471
6–8	720–726	624–631
9–12	796–798	685–692

#### *IV.1.3.3 Raters*

KELPA constructed responses were scored by qualified educators. DTCs assigned qualified educators within a school district to score KELPA CR items in speaking and writing. Students assigned to receive two ratings were rated by DTC-assigned educators who were different from the raters who rated the primary score. The first and second ratings are considered interchangeable in score quality, since scorers were expected to receive the same level of training and be familiar with the scoring rubrics. Refer to [Section II.3.1 Test-Administrator and Scorer Training](#) and [Section IV.3.1.2 Educator Scoring](#) in the 2020 *KELPA Technical Manual* for details about rater training and assignment.

#### IV.1.3.4 Interrater Agreement

##### IV.1.3.4.1 Methods

*Agreement* measures how frequently two raters assign the same rating (Graham et al., 2012). The percentage of items on which raters agree exactly is referred to as *exact agreement*; the percentage of items on which raters agree either exactly or within one point of each other is referred to as *adjacent agreement*. In general, an exact agreement level of 75% or above is acceptable for most fields, and exact-plus-adjacent agreements should be 90% or above (Graham et al., 2012). *Kappa* originally measured the agreement between two raters on a two-level (i.e., pass vs. fail) rating scale, but kappa can also measure agreement when three or more performance levels are used. *Weighted kappa* distinguishes between the number of ratings falling within one performance level and the numbers of ratings that differ by two or more performance levels (Graham et al., 2012). The *quadratic-weighted kappa* is calculated using expected scores and predicted scores, and it measures the agreement between two ratings; the value typically ranges from 0 (i.e., random agreement between raters) to 1 (i.e., complete agreement between raters). When there is less agreement between raters than expected by chance, the value may go below 0. For example, suppose rater A assigns a sample of  $n$  subjects across  $m$  categories of a categorical scale, and suppose rater B independently does the same thing. Equation IV-2 shows how the mean observed degree of disagreement is calculated, and Equation IV-3 shows how the mean degree of disagreement expected by chance (i.e., expected if A and B assign subjects randomly in accordance with their respective base rates) is calculated (Fleiss & Cohen, 1973):

$$\bar{D}_o = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^m n_{ij} v_{ij}, \quad (\text{IV-2})$$

$$\bar{D}_e = \frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^m n_{i \cdot} n_{\cdot j} v_{ij}, \quad (\text{IV-3})$$

where  $n_{ij}$  denotes the number of subjects assigned to category  $i$  by rater A and to category  $j$  by rater B,  $n_{i \cdot}$  denotes the total number of subjects assigned to category  $i$  by rater A, and  $n_{\cdot j}$  denotes the total number of subjects assigned to category  $j$  by rater B;  $v_{ij}$  denotes the disagreement weight associated with categories  $i$  and  $j$ .

When  $v_{ij} = 0$ , it reflects no disagreement when a subject is assigned to category  $i$  by both raters; when  $v_{ij} > 0$ , for  $i \neq j$ , it reflects some degree of disagreement when a subject is assigned to various categories by the two raters. Quadratic-weighted kappa is then defined by Equation IV-4 (Fleiss & Cohen, 1973):

$$k_w = \frac{\bar{D}_e - \bar{D}_o}{\bar{D}_e}. \quad (\text{IV-4})$$

It is a special case of weighted kappa when  $v_{ij} = 1$  for all  $i \neq j$ . The quadratic weight emphasizes the importance of near-disagreement and drops quickly when there are two or more category differences. A kappa value greater than .75 indicates excellent agreement, a value less than .40 indicates poor agreement, and any value between .40 and .75 indicates good agreement (Cohen, 1968).

##### IV.1.3.4.2 Results

Table IV-7 summarizes rater agreement for writing items. For writing responses, the average percentage of exact agreement across items within a grade or grade band—both overall (i.e., mean percentage of agreement on all responses, regardless of the scoring method applied) and for the individual scoring

method—ranged from 58% (grade band 6–8) to 84% (grade 1). The average percentage of exact-plus-adjacent agreement across items within a grade or grade band—both overall and for the individual scoring method—was 96% or above.

*Table IV-7. Rater Agreement on Writing Items Scored Using the Individual Scoring Method by Grade or Grade Band*

Grade or grade band	Mean exact agreement across items (%)		Mean exact-plus-adjacent agreement across items (%)	
	Overall	Individual scoring	Overall	Individual scoring
K	79	78	97	97
1	84	84	98	98
2–3	76	76	97	97
4–5	70	70	98	98
6–8	58	58	96	96
9–12	61	61	96	96

Table IV-8 summarizes agreement for speaking items. For speaking responses, the average percentage of exact agreement across items within a grade or grade band—for overall (i.e., mean percentage of agreement on all responses regardless of scoring method applied), the individual scoring method, and the combination of individual and deferred scoring methods—ranged from 63% (kindergarten) to 72% (grade band 2–3). The average percentage of exact-plus-adjacent agreement across items within a grade or grade band—for overall, the individual scoring method, and the combination of individual and deferred scoring methods—was 95% or greater.

*Table IV-8. Rater Agreement on Speaking Items*

Grade or grade band	Mean exact agreement across items (%)			Sum of mean exact plus adjacent agreement across items (%)		
	Overall	Individual scoring	Individual + deferred	Overall	Individual scoring	Individual + deferred
K	63	63	64	95	95	96
1	64	64	63	95	95	95
2–3	72	71	70	98	98	98
4–5	71	71	71	98	98	98
6–8	67	67	67	96	96	96
9–12	68	68	68	95	95	94

Table IV-9 shows the classifications of quadratic-weighted kappa values of KELPA CR items. To be consistent with Table IV-5, Table IV-6, Table IV-7, and Table IV-8, the number of items with excellent or good agreement reported in Table IV-9 is based on responses scored using the individual scoring method for writing items and the combination of individual and deferred scoring methods for speaking items. Quadratic-kappa results show that all items had good to excellent agreement. Excellent agreement was found for responses to writing items in grades 1–3. For both speaking and writing, lower grades (i.e., kindergarten through grade 3) had better agreement than higher grades.

Table IV-9. Summary of Quadratic Kappa Classifications

Grade or grade band	No. of items (% of domain items)			
	Writing		Speaking	
	Excellent agreement	Good agreement	Excellent agreement	Good agreement
K	3(75)	1(25)	8(80)	2(20)
1	4(100)	0(0)	1(10)	9(90)
2–3	4(100)	0(0)	8(80)	2(20)
4–5	2(50)	2(50)	2(20)	8(80)
6–8	0(0)	3(100)	4(40)	6(60)
9–12	1(33)	2(67)	9(90)	1(10)

#### IV.1.3.4.3 Summary

Individual scoring was the dominant scoring method for both writing and speaking items in 2023. Individual scoring paired with deferred scoring was the dominant scoring method for speaking. The average percentage of exact agreement between two raters across items within a grade or grade band ranged from 58% to 84% for writing responses and from 63% to 72% for speaking responses. The average percentages of exact-plus-adjacent agreement across items within a grade or grade band were 96% or greater for writing responses and 95% or greater for speaking responses. Statistics for the quadratic-weighted kappa show that, for writing responses, raters had excellent agreement on all items in grade 1 and grade band 2–3, and a mixture of good to excellent agreement on other grades' items. For speaking responses, raters had a mixture of good to excellent agreement on items in all grades. The degree of rater agreements based on agreement rates and quadratic kappa appears to point to similar conclusions. Both kindergarten writing and speaking seemed to have slightly lower rater agreements than other grades.

## IV.2 Scoring and Scaling

This section provides test-result summaries for the 2023 administration. For information about the procedures for scoring individual items, scoring the test, scaling, and specific quality-control process followed by AAI and Agile Technology Solutions to ensure the accuracy of scoring results, refer to [Section IV.3.5 Quality-Control Checks](#) of the *2020 KELPA Technical Manual*.

### IV.2.1 Operational Test Results

The number of students who took KELPA in 2023, along with a summary of their demographic characteristics, is provided in this section. Operational test results present the summary statistics of test scores, which show the distribution of students' test scores. Statistics for test scores by domain for the entire population and for different student groups were calculated and are summarized below. Also, the percentages of students in each performance level are included in this section.

#### IV.2.1.1 Test-Enrollment Data

All students who are identified as ELs must take KELPA. For students registered for the first time in K–12 schools in Kansas, a home-language survey is used to determine whether a student is a potential EL.

A student who is identified by the home-language survey as a potential EL is required to take a Kansas State Department of Education (KSDE)-approved EL screener to determine whether KELPA is required. A potential EL who does not pass the screener is considered an EL and will take KELPA in the spring. Students who scored as proficient on KELPA in 2023 are not required to take KELPA again in the next school year.

KELPA was administered in the four domains: listening, speaking, reading, and writing. Students who took the tests were in grades K–12. Students who viewed a listening or reading test, even if they did not answer any questions, are categorized as having taken the domain test. For the writing and speaking tests, students are categorized as having taken the domain test if a teacher has scored the tests, even if students did not answer any items. Students who took at least one domain test received a score report and will be considered to have participated in the test. Table IV-20 in Section IV.2.2.1 Comparison of Enrollment in the current manual presents the number and percentage of enrolled students who were tested in each grade for KELPA administrations from 2021 to 2023. The participation rate or tested rate for 2023 KELPA, computed as number of students tested divided by number of students enrolled, ranged from 89% to 99%, with the lowest participation rates in high school grades.

The participation rates for the 10 State Board of Education (SBOE) districts in 2023 are presented in Table IV-10 by grade or grade band. Kansas has 286 school districts that are grouped into 10 SBOE districts. The participation rates (i.e., tested rates) ranged from 93% (SBOE districts 7 and 8 in grade band 9–12) to 100% (SBOE districts 6 and 9 in grade 1 and SBOE district 9 in grade band 6–8). The tested rates were lower in grade band 9–12 across all SBOE districts than in other grades and grade bands. The two largest school districts were the Kansas City, Kansas Public Schools district (part of SBOE district 1, whose average tested rate was 99% across grades and grade bands) and the Wichita Public Schools district (part of SBOE district 7, whose average tested rate was 98% across grades and grade bands). Both school districts were in SBOE districts that had remarkably high participation rates in elementary and middle schools but decreased participation rates in high schools. The trend of participation rates in high schools in these two SBOE districts was consistent with the dramatic enrollment drop from 2021 to 2022, which then bounced back in 2023 in grades 10–12, as reported in Table IV-20. This indicates that the two largest school districts have undergone a significant impact from the pandemic but have experienced a quick recovery on both enrollment and participation rates.

Table IV-10. 2023 KELPA Participation Rates by State Board of Education (SBOE) District and Grade or Grade Band

SBOE district	Kindergarten		Grade 1		Grade band 2–3		Grade band 4–5		Grade band 6–8		Grade band 9–12	
	Enrolled students	Tested students	Enrolled students	Tested students	Enrolled students	Tested students	Enrolled students	Tested students	Enrolled students	Tested students	Enrolled students	Tested students
	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)
1	1,160	100 <sup>a</sup>	1,300	100 <sup>a</sup>	2,271	100 <sup>a</sup>	1,883	100 <sup>a</sup>	2,213	99	2,434	96
2	1,613	100 <sup>a</sup>	1,717	100 <sup>a</sup>	2,862	100 <sup>a</sup>	2,303	100 <sup>a</sup>	2,697	99	2,965	96
3	807	99	757	99	1,210	99	890	100 <sup>a</sup>	1,052	99	1,139	97
4	1,698	100 <sup>a</sup>	1,783	100 <sup>a</sup>	3,066	100 <sup>a</sup>	2,513	100 <sup>a</sup>	2,943	99	3,322	96
5	984	100 <sup>a</sup>	997	100 <sup>a</sup>	1,779	100 <sup>a</sup>	1,511	100 <sup>a</sup>	1,829	100 <sup>a</sup>	2,091	98
6	419	99	413	100	761	99	598	99	691	98	826	97
7	997	99	1,056	99	1,974	99	1,601	98	1,967	98	2,318	93
8	788	99	891	99	1,700	99	1,380	98	1,721	98	2,066	93
9	172	99	126	100	215	99	209	100 <sup>a</sup>	189	100	248	98
10	873	99	968	99	1,857	99	1,529	98	1,846	98	2,200	94

<sup>a</sup> Calculated as 100% with rounding.

For all tested ELs, Table IV-11 shows the percentage of students in each demographic group by grade.<sup>5</sup> The groups include race, ethnicity, disability status, and gender. The percentage of students in each student group was remarkably similar across grades, except there were more American Indian students in higher grades and fewer White students in higher grades. The majority race group was White, the majority ethnicity group was Hispanic, and there were about equal percentages of male and female students, with slightly more male students in each grade.

<sup>5</sup> 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 IV-11. Percentage of Tested Students by Demographic Characteristic and Grade

Characteristic	Grade (%)												
	K (n = 4,579)	1 (n = 4,742)	2 (n = 4,405)	3 (n = 3,915)	4 (n = 3,585)	5 (n = 3,161)	6 (n = 2,733)	7 (n = 2,627)	8 (n = 2,564)	9 (n = 2,683)	10 (n = 2,443)	11 (n = 1,913)	12 (n = 1,530)
Race													
Black	4.8	4.2	4.9	5	4.6	4.7	5	4.9	4.9	5.6	6.1	5.5	6.6
American Indian	5.9	6.3	6.2	7	8.2	7.1	8.2	8.8	9.8	9.6	10.7	11.9	15.2
Asian	10.6	11.7	9.7	9.2	9	8	8	8	7.1	6.6	7.6	6.6	7.2
NHPI	1.4	1.2	1	1.6	1.4	1.3	1.5	1.7	1.4	1.1	1.1	1.1	0.9
White	77.2	76.6	78.2	77.1	76.8	78.9	77.3	76.6	76.8	77	74.5	74.8	70.1
Hispanic													
Yes	77.9	78	80.1	80.9	82.1	82.9	82.3	82.8	83.8	83.6	82.7	84.4	82.3
No	22.1	22	19.9	19.1	17.9	17.1	17.7	17.2	16.2	16.4	17.3	15.6	17.7
SWD													
Yes	11.1	11.7	12.6	14.5	15.8	19.5	19.4	19.7	20.7	18.8	16.1	17.9	15.3
No	88.9	88.3	87.4	85.5	84.2	80.5	80.6	80.3	79.3	81.2	83.9	82.1	84.7
Gender													
Female	46.7	49.1	46.7	45.1	47.1	45.7	43.2	44.5	43.5	42.9	43.6	43.2	43.7
Male	53.3	50.9	53.3	54.9	52.9	54.3	56.8	55.5	56.5	57.1	56.4	56.8	56.3

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

#### IV.2.1.2 Test Results for All Students

Summaries of scale scores by grade and domain are presented in Table IV-12, Table IV-13, Table IV-14, and Table IV-15. As the tables show, the minimum and maximum values were within the lowest obtainable scale score (LOSS; i.e., 0) and the highest obtainable scale score (HOSS; i.e., 1,000), respectively. Although grades and domains use the same scale score with the same LOSS and HOSS, the assessments are not linked across domains and grades. Thus, the same score has different meanings across domains and grades, and scores across domains and grades should not be compared. In the summary tables below, the 10th, 25th, 50th, 75th, and 90th percentiles were provided as  $P_{10}$ ,  $P_{25}$ ,  $P_{50}$ ,  $P_{75}$ , and  $P_{90}$ , respectively. The differences between (a)  $P_{50}$  and  $P_{25}$  and (b)  $P_{75}$  and  $P_{50}$ , respectively, indicate the shape of score distributions; the larger of the two differences indicates the direction of any skewness in the distribution (i.e., a negative skew when the first difference is larger and a positive skew when the second difference is larger). If the two differences match, the distribution is symmetric. For the listening test, the



distribution of scale scores was negatively skewed in grades 2–4, 11, and 12, and positively skewed in other grades. For the speaking test, the distribution of scale scores was positively skewed in grades 2–4 and 8; distributions for other grades were skewed negatively. For the reading test, the distribution of scale scores was negatively skewed in grades 8, 10, and 12, and positively skewed in other grades. For the writing test, the distribution of scale scores was positively skewed in grades 1, 5, and 7, and negatively skewed in other grades.

*Table IV-12. Scale-Score Descriptive Statistics by Grade for Listening*

Grade	M	SD	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	526.86	169.99	0	354	421	492	589	695	1000
1	499.61	141.31	0	343	421	480	573	626	1000
2	492.80	167.16	0	328	404	475	541	605	1000
3	571.49	209.70	0	365	453	541	605	1000	1000
4	501.23	169.37	0	326	411	491	535	611	1000
5	544.58	197.51	0	349	432	491	611	1000	1000
6	474.89	116.04	0	335	414	478	552	615	1000
7	503.89	137.90	0	347	432	478	552	615	1000
8	529.93	160.08	66	335	432	510	615	725	1000
9	464.02	143.79	139	303	371	455	547	622	1000
10	480.25	160.26	0	315	371	455	547	622	1000
11	505.15	168.54	200	327	394	477	547	622	1000
12	516.54	177.87	139	327	394	506	547	622	1000

*Note.* M = Mean, SD = Standard Deviation. P<sub>10</sub>, P<sub>25</sub>, P<sub>50</sub>, P<sub>75</sub>, and P<sub>90</sub> are the 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

*Table IV-13. Scale-Score Descriptive Statistics by Grade for Speaking*

Grade	M	SD	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	484.18	163.9	0	311	434	515	580	634	1000
1	528.18	186.38	0	366	448	526	576	640	1000
2	519.46	179.98	0	353	446	500	575	616	1000
3	561.57	201.14	0	386	472	531	616	1000	1000
4	548.14	215.61	0	366	447	502	577	1000	1000
5	566.24	227.11	0	366	460	520	577	1000	1000
6	504.22	197.16	0	337	430	496	536	583	1000
7	521.68	217.71	0	346	430	496	555	1000	1000
8	541.09	241.31	0	337	440	508	583	1000	1000
9	493.93	265.72	0	0	399	485	535	1000	1000
10	515.71	272.13	0	0	411	493	556	1000	1000
11	536.15	295.22	0	0	405	502	556	1000	1000
12	523.7	307.52	0	0	405	502	556	1000	1000

*Note.* M = Mean, SD = Standard Deviation. P<sub>10</sub>, P<sub>25</sub>, P<sub>50</sub>, P<sub>75</sub>, and P<sub>90</sub> are the 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

*Table IV-14. Scale-Score Descriptive Statistics by Grade for Reading*

Grade	M	SD	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	485.81	127.82	0	363	399	463	552	618	1000
1	481.04	126.8	0	369	393	451	548	648	1000
2	463.29	123.01	0	347	377	441	516	606	1000
3	528.67	160.69	0	362	416	498	606	673	1000
4	478.79	131.17	0	329	388	465	557	665	1000
5	511.85	141.05	0	344	404	491	602	665	1000
6	469.68	112.58	0	336	390	463	541	579	1000
7	497.79	127.6	0	355	407	485	579	628	1000
8	516.29	135.86	113	355	424	511	579	699	1000
9	451.13	107.46	0	338	377	439	521	594	1000
10	460.33	112.2	0	338	377	454	521	594	1000
11	483.24	118.67	0	338	393	469	566	631	1000
12	491.71	128.14	0	338	393	485	566	631	1000

*Note.* M = Mean, SD = Standard Deviation. P<sub>10</sub>, P<sub>25</sub>, P<sub>50</sub>, P<sub>75</sub>, and P<sub>90</sub> are the 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

*Table IV-15. Scale-Score Descriptive Statistics by Grade for Writing*

Grade	M	SD	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	497.53	145.75	0	342	426	499	554	637	1000
1	498.97	158.99	0	336	400	464	588	691	1000
2	461.22	127.56	0	298	381	465	548	622	1000
3	510.87	135.03	0	342	434	523	580	687	1000
4	476.72	125.35	0	319	400	479	532	600	1000
5	511.68	139.22	0	335	437	504	600	649	1000
6	489.55	146.7	0	327	410	496	557	607	1000
7	507.33	155.41	0	327	428	496	596	652	1000
8	536.76	179.41	0	340	428	525	596	652	1000
9	434.75	134.03	0	273	375	448	511	555	1000
10	447.35	131.65	0	300	375	468	532	585	1000
11	469.31	145.83	0	321	393	490	532	632	1000
12	470.15	161.61	0	300	393	490	555	632	1000

*Note.* M = Mean, SD = Standard Deviation. P<sub>10</sub>, P<sub>25</sub>, P<sub>50</sub>, P<sub>75</sub>, and P<sub>90</sub> are the 10th, 25th, 50th, 75th, and 90th percentiles, respectively.

The proportion of students in each performance level (i.e., levels 1 through 4) is shown by domain and grade in Figure IV-1, Figure IV-2, Figure IV-3, and Figure IV-4. Students must obtain level 4 in each of the four domains to be considered proficient overall. The percentage of students in level 4 ranged from 27% (grade 1) to 70% (grade 3) across grades for listening, from 20% (kindergarten) to 55% (grades 3 and 4) across grades for speaking, from 10% (kindergarten) to 40% (grade 2) across grades for reading, and from 9% (kindergarten) to 37% (grade 5) across grades for writing.

Figure IV-1. 2023 Performance-Level Results for Listening

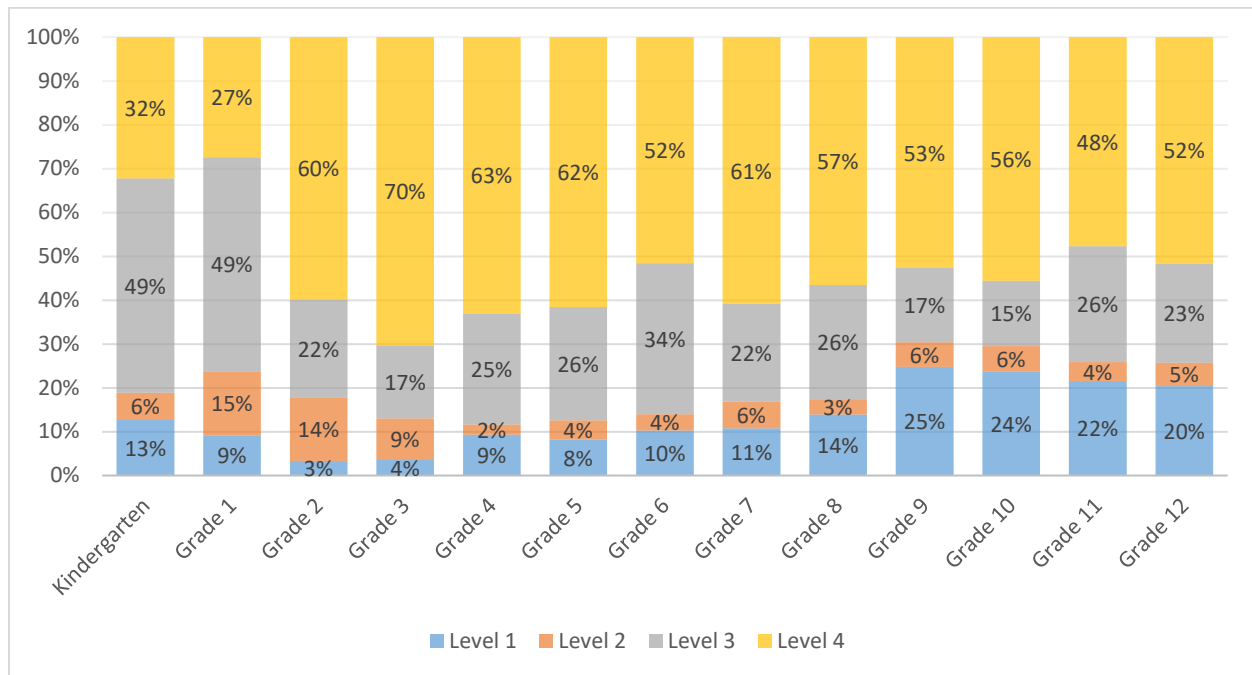


Figure IV-2. 2023 Performance-Level Results for Speaking

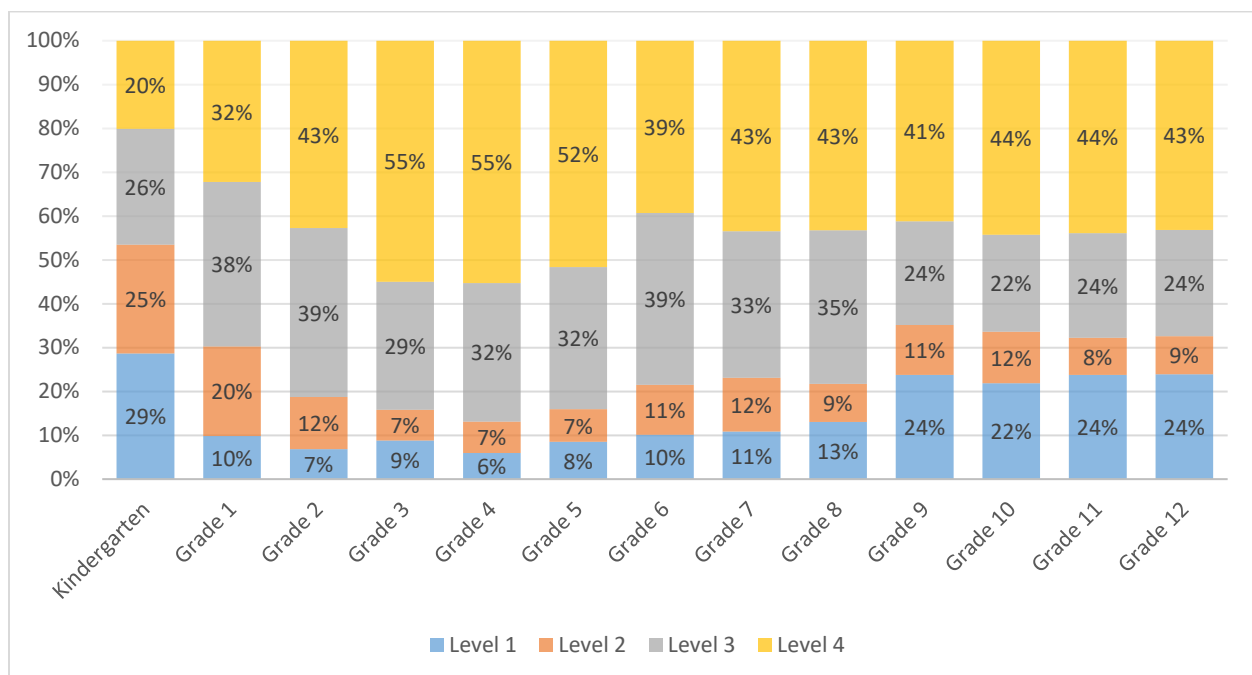


Figure IV-3. 2023 Performance-Level Results for Reading

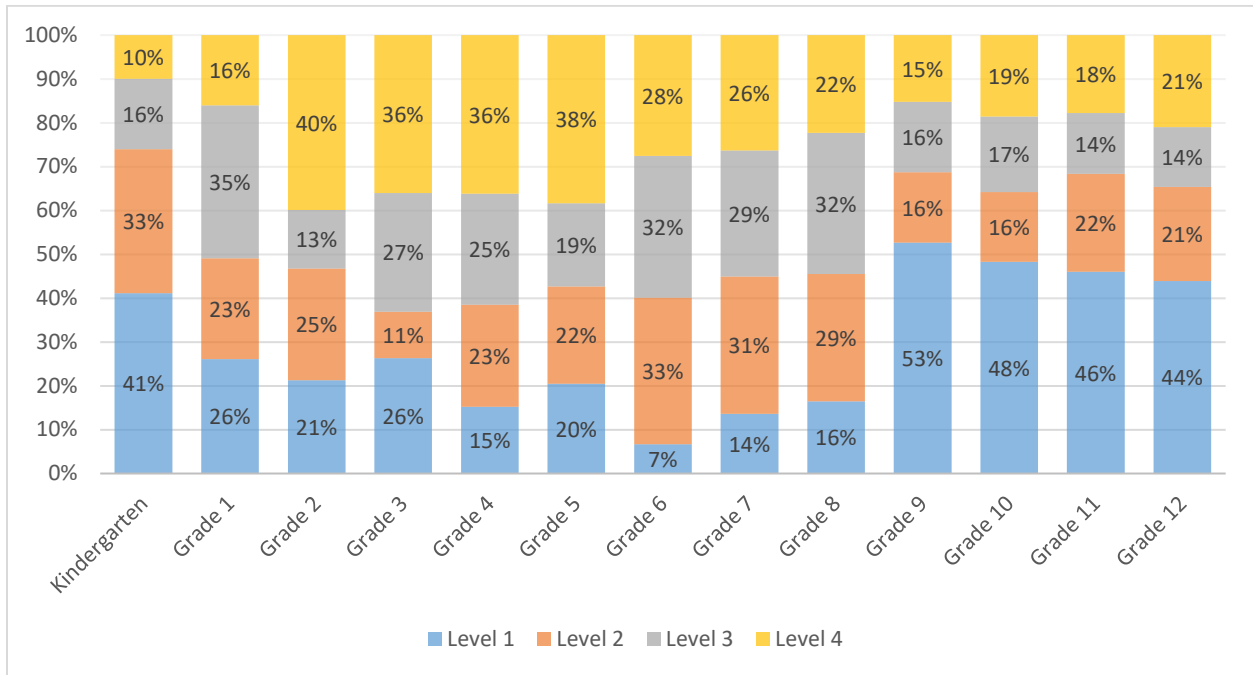
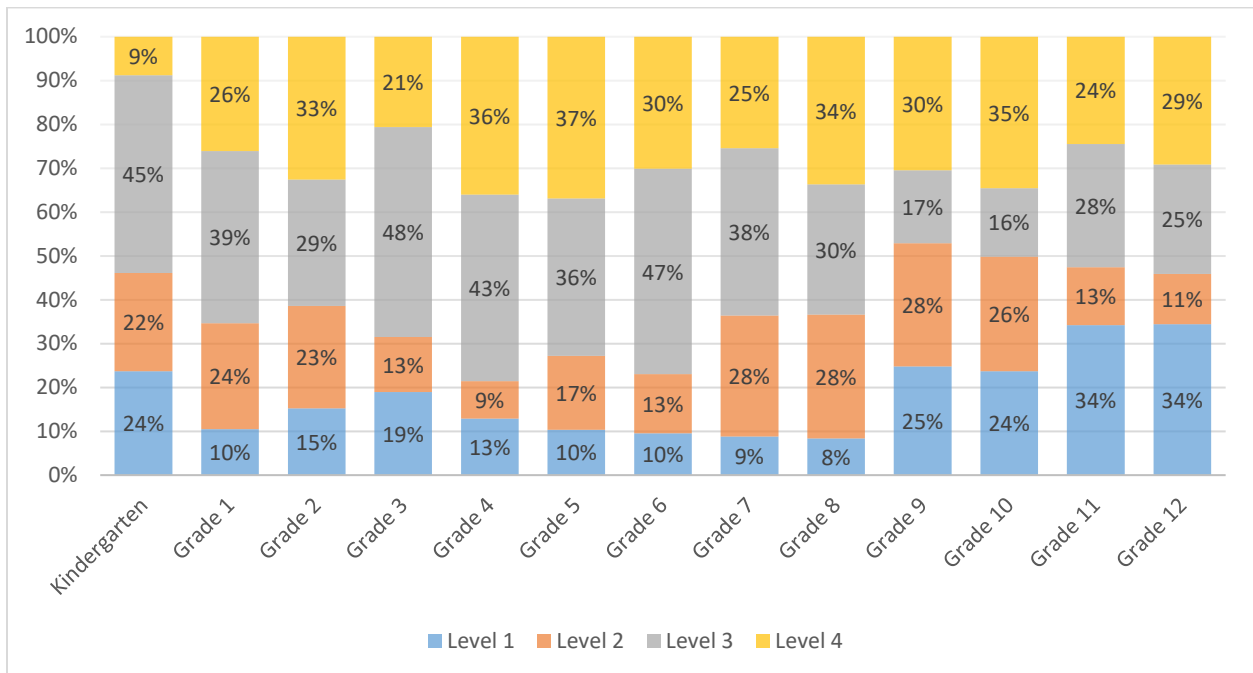


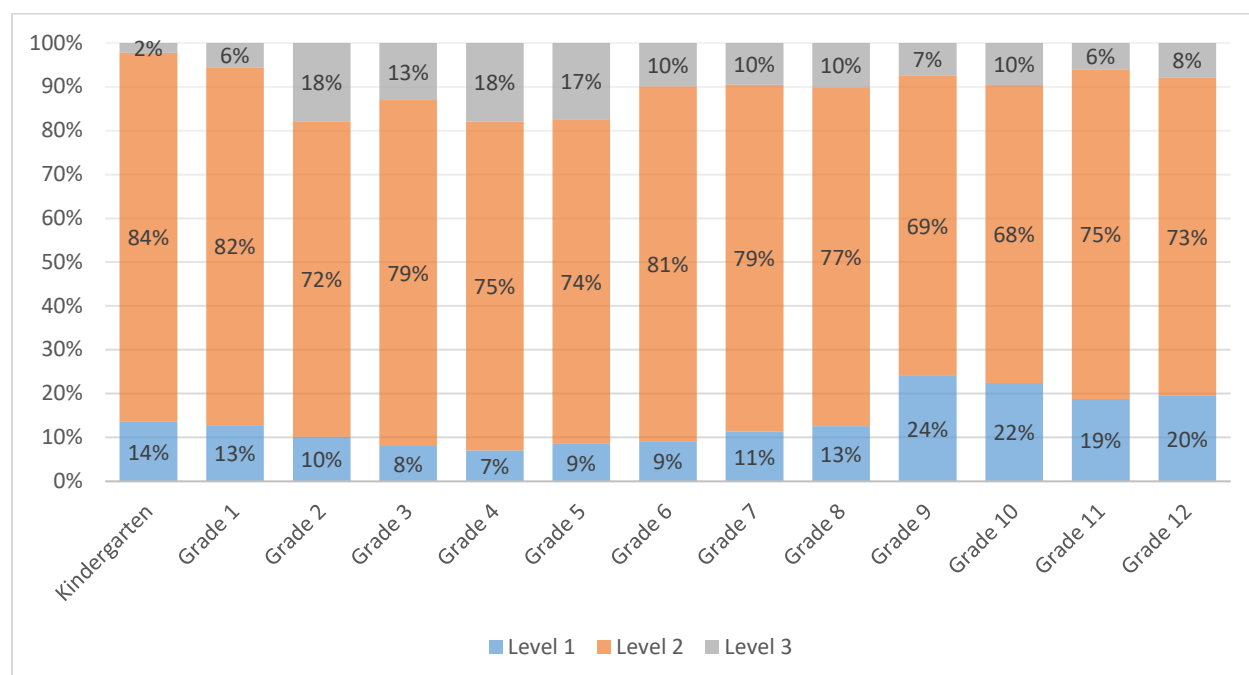
Figure IV-4. 2023 Performance-Level Results for Writing



The overall proficiency levels are determined from the four domain performance levels. When students are categorized as level 4 on all four domain tests, the overall proficiency level is level 3 (i.e., proficient). When students are at either level 1 or level 2 on all four domain tests, the overall proficiency level is

level 1 (i.e., not proficient). Students not classified as proficiency level 3 or level 1 are at level 2 (i.e., nearly proficient). The overall proficiency levels in 2023 are presented in Figure IV-5. Results indicate that most students were categorized as level 2; the percentages ranged from 69% (grade 9) to 84% (kindergarten). Overall, the proficiency rates ranged from 2% (kindergarten) to 18% (grades 2 and 4). Kindergarten had lower percentages of students in level 3 compared to other grades, which is expected and consistent with results in previous years, given that students in early grades have had little exposure to formal instruction or English for speakers of other languages services.

*Figure IV-5. Overall Performance-Level Results (2023 Administration)*



#### IV.2.1.3 Student-Group Test Results

Summaries of average scale scores by demographic groups<sup>6</sup> are presented in Table IV-16, Table IV-17, Table IV-18, and Table IV-19. For group sample sizes, refer to Table IV-11. In most grades and domains, Asian students had the highest mean scores. For the listening test, American Indian (AI) students had the highest mean scores in grades 4, 5 (tied with White), 7, and 10. Native Hawaiian and Pacific Islander (NHPI) students had the highest mean scores in grade 6. For the speaking test, Black students at grades 2 and 7; NHPI students at grades 4, 5, and 10; and White students at grade 3 had the highest mean scores. For the reading test, AI students at grade 10 and NHPI students at grades 4 and 6 had the highest mean scores. For the writing test, AI students at grade 6 had the highest mean scores. Across all domains, the mean scores of non-Hispanic students were higher than those of Hispanic students in most grades and were slightly lower in some grades (i.e., grades 3, 4, 5, and 7 in listening; grades 3 and 5 in speaking). Across all domains and grades, the mean scores of students without a disability were higher than those of students with a disability, except for speaking in grade 9. For speaking and writing tests,

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

the mean scores of female students were higher than those of male students in all grades. The mean scores of female students were higher than those of male students in most grades, except for grades 4, 5, and 9 in listening and for grades 5 (tied), 9, 10, and 11 in reading. These findings are similar to 2022 findings. Even when a test is carefully constructed with many considerations of fairness, differences may exist among student groups due to achievement gaps. Trend data comparing both the overall test results and results in each domain from 2021 to 2023 are provided in the next subsection.

Table IV-16. Demographic Group Scale-Score Descriptive Statistics by Grade for Listening

Group	K		1		2		3		4		5		6		7		8		9		10		11		12		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Race																											
AI	503	152	493	135	481	170	578	211	505	172	546	194	463	117	509	148	540	163	461	135	492	167	481	153	493	173	
Asian	552	190	509	149	514	186	579	234	503	190	533	207	478	124	496	141	542	175	489	153	491	168	533	167	538	185	
Black	500	165	489	164	494	187	532	186	485	164	523	203	472	104	497	140	502	162	461	165	483	177	514	166	512	181	
NHPI	497	134	471	108	443	120	544	214	482	118	509	168	484	128	470	124	469	95	422	70	479	145	486	160	489	114	
White	527	170	499	139	492	164	573	208	501	168	546	196	475	116	505	137	530	158	462	142	478	157	507	172	519	177	
Hispanic																											
Yes	522	165	498	140	490	164	573	208	502	167	546	195	474	114	505	137	529	158	462	142	477	158	501	168	510	174	
No	545	187	504	146	503	180	564	219	500	179	538	211	480	125	498	141	535	170	477	155	497	170	526	170	546	193	
SWD																											
Yes	468	168	439	129	440	160	507	188	448	150	488	161	435	92	481	125	506	129	447	95	461	132	485	142	492	150	
No	534	169	508	141	500	167	582	211	511	171	558	203	484	119	509	140	536	167	468	153	484	165	510	173	521	182	
Gender																											
Female	542	169	514	143	503	167	580	209	498	160	534	188	482	114	512	145	537	163	462	140	489	167	508	168	528	172	
Male	513	169	486	138	484	167	564	210	504	177	553	205	469	118	497	132	525	158	466	146	474	154	503	169	508	182	

Note. M = Mean, SD = Standard Deviation. AI = American Indian; NHPI = Native Hawaiian and Pacific Islander; SWD = students with disabilities.

Table IV-17. Demographic Group Scale-Score Descriptive Statistics by Grade for Speaking

Group	K		1		2		3		4		5		6		7		8		9		10		11		12		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Race																											
AI	467	161	501	163	502	165	541	202	510	172	549	210	492	224	528	257	544	240	513	248	517	263	554	294	502	304	
Asian	508	188	548	194	529	190	561	205	561	224	563	235	520	212	518	216	562	219	527	266	500	259	576	278	583	278	
Black	498	167	527	199	550	207	533	187	550	216	571	249	483	148	535	219	491	246	495	242	508	290	563	246	560	292	
NHPI	484	129	539	153	514	154	555	189	568	207	618	238	516	170	511	278	532	199	460	282	558	264	550	351	569	349	
White	480	162	527	186	517	178	565	202	549	219	567	226	504	196	520	211	542	243	489	269	517	273	528	299	519	313	
Hispanic																											
Yes	479	161	522	183	515	177	563	201	548	214	568	227	503	195	520	218	538	241	490	269	514	272	529	298	512	309	
No	504	174	549	196	536	191	555	203	550	223	557	229	510	206	530	215	558	241	515	250	526	272	576	277	578	294	
SWD																											
Yes	394	195	452	169	459	168	508	163	504	184	532	196	475	180	505	214	531	226	504	223	515	265	496	286	487	295	
No	495	156	538	186	528	180	571	206	556	220	575	233	511	201	526	219	544	245	492	275	516	274	545	297	530	309	
Gender																											
Female	499	165	547	189	531	182	584	214	567	226	584	239	524	209	530	226	556	249	496	269	531	283	549	298	547	314	
Male	471	162	510	182	509	178	543	188	531	204	551	215	489	186	515	211	529	234	492	264	504	263	526	293	505	301	

Note. M = Mean, SD = Standard Deviation. AI = American Indian; NHPI = Native Hawaiian and Pacific Islander; SWD = students with disabilities.



Table IV-18. Demographic Group Scale-Score Descriptive Statistics by Grade for Reading

Group	K		1		2		3		4		5		6		7		8		9		10		11		12		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Race																											
AI	465	100	464	115	450	112	511	150	461	119	498	131	459	122	493	118	524	131	438	88	470	114	474	117	476	127	
Asian	566	172	553	160	499	133	561	187	508	158	519	153	478	114	510	142	545	158	479	109	464	115	503	141	500	112	
Black	498	155	494	122	469	125	512	161	459	129	513	140	467	107	490	132	486	145	429	112	438	107	476	120	459	130	
NHPI	469	108	447	90	458	117	495	119	525	157	504	149	487	118	477	138	521	135	423	79	450	88	473	93	466	106	
White	476	118	471	119	460	121	527	158	477	129	512	140	469	112	497	125	514	132	452	108	460	111	484	117	498	129	
Hispanic																											
Yes	471	111	468	117	457	120	525	157	476	127	511	139	469	112	497	126	514	133	449	105	459	111	481	116	490	127	
No	538	166	527	148	488	131	543	175	491	147	517	150	475	115	501	134	526	149	463	120	465	116	494	132	501	135	
SWD																											
Yes	442	137	434	108	412	97	453	127	416	111	451	109	422	90	453	106	471	116	432	82	445	102	469	115	466	113	
No	491	126	487	128	471	125	541	162	491	131	527	144	481	114	509	130	528	138	456	112	463	114	486	119	496	130	
Gender																											
Female	490	126	482	122	469	120	536	161	480	132	512	138	473	105	498	123	520	133	443	97	458	105	480	109	495	114	
Male	482	130	480	132	458	125	523	160	477	131	512	143	467	118	497	131	513	138	457	114	462	118	485	126	489	138	

Note. M = Mean, SD = Standard Deviation. AI = American Indian; NHPI = Native Hawaiian and Pacific Islander; SWD = students with disabilities.

Table IV-19. Demographic Group Scale-Score Descriptive Statistics by Grade for Writing

Group	K		1		2		3		4		5		6		7		8		9		10		11		12		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Race																											
AI	477	120	488	160	442	109	498	136	466	127	491	130	474	149	498	144	553	185	441	111	455	122	466	148	455	153	
Asian	563	182	566	191	508	147	539	158	510	153	545	169	509	176	538	179	559	192	477	145	458	133	524	160	510	154	
Black	505	165	502	153	463	128	499	143	457	118	501	126	469	137	505	205	500	182	396	139	420	126	451	118	454	117	
NHPI	493	139	484	150	470	111	508	126	509	117	517	150	539	164	487	162	549	167	382	165	443	93	481	179	450	168	
White	489	139	489	151	456	126	509	132	473	122	510	136	489	143	506	151	535	177	434	134	448	132	467	145	472	169	
Hispanic																											
Yes	485	134	486	149	453	123	506	129	474	121	508	135	489	142	503	149	535	177	431	130	446	131	464	145	464	163	
No	544	175	544	183	495	141	532	155	489	143	530	159	495	166	527	184	544	190	453	151	455	134	496	150	500	151	
SWD																											
Yes	431	170	424	140	406	126	440	118	407	106	450	113	426	118	456	117	502	147	431	97	439	115	448	138	437	160	
No	506	140	509	159	469	126	523	134	490	124	527	141	505	149	520	161	546	186	436	141	449	135	474	147	476	161	
Gender																											
Female	505	142	509	154	472	123	522	136	489	128	524	140	511	149	525	170	565	194	448	139	470	131	484	156	497	167	
Male	491	148	489	163	452	131	502	134	466	122	501	138	473	143	493	141	515	164	425	129	430	130	458	137	449	154	

Note. M = Mean, SD = Standard Deviation. AI = American Indian; NHPI = Native Hawaiian and Pacific Islander; SWD = students with disabilities.

## IV.2.2 Trend Data

The 2023 KELPA administration was the fifth administration of the new KELPA aligned with the [2018 Standards](#). The next subsections present changes in enrollment data and performance-level distributions from 2021 to 2023.

### IV.2.2.1 Comparison of Enrollment

Due to the impact of the COVID-19 pandemic on the 2020–2021 academic school year, the enrollment and test-participation rates greatly decreased in each grade but started to pick up and continued to increase in the past three years (see Table IV-20). For the 2023 administration, 41,528 students were enrolled and 40,880 students tested; the overall participation rate was 98%. Participation rates across grades ranged from 89% (grade 12) to 99% (kindergarten through grade 8). Compared to the 2022 administration, the enrollments in 2023 continued to increase for grades 1–6, 8, and 10 but also decreased for kindergarten and grades 7, 9, 11, and 12. On average, the enrollments in 2023 increased by 2% compared to the 2022 administration.

Table IV-20. Number and Percentage of Enrolled and Tested Students by Grade: 2020 Through 2022

Grade	2021			2022				2023			
	No. enrolled	No. tested	Participation %	No. enrolled	No. tested	Participation %	% Enrollment change (2021 to 2022)	No. enrolled	No. tested	Participation %	% Enrollment change (2022 to 2023)
<b>K</b>	4,305	4,090	95	4,638	4,597	99	8	4,603	4,579	99	−1
<b>1</b>	4,434	4,212	95	4,471	4,436	99	1	4,767	4,742	99	7
<b>2</b>	4,336	4,119	95	4,376	4,342	99	1	4,445	4,405	99	2
<b>3</b>	3,926	3,730	95	3,929	3,884	99	0	3,949	3,915	99	1
<b>4</b>	3,536	3,359	95	3,623	3,583	99	2	3,628	3,585	99	0
<b>5</b>	3,041	2,889	95	3,114	3,061	98	2	3,194	3,161	99	3
<b>6</b>	2,724	2,452	90	2,692	2,639	98	−1	2,770	2,733	99	3
<b>7</b>	2,538	2,310	91	2,684	2,619	98	6	2,665	2,627	99	−1
<b>8</b>	2,480	2,207	89	2,424	2,387	98	−2	2,596	2,564	99	7
<b>9</b>	2,551	2,092	82	2,844	2,736	96	11	2,724	2,683	98	−4
<b>10</b>	2,495	1,996	80	2,205	2,093	95	−12	2,487	2,443	98	13
<b>11</b>	2,373	1,780	75	2,003	1,878	94	−16	1,973	1,913	97	−1
<b>12</b>	2,094	1,361	65	1,823	1,510	83	−13	1,727	1,530	89	−5
<b>Total</b>	<b>40,834</b>	<b>36,597</b>	<b>90</b>	<b>40,826</b>	<b>39,765</b>	<b>97</b>	<b>0</b>	<b>41,528</b>	<b>40,880</b>	<b>98</b>	<b>2</b>

Note. Positive values indicate a percentage increase; negative values indicate a percentage of decrease.

#### IV.2.2.2 Comparison of Performance-Level Results

Figure IV-6, Figure IV-7, Figure IV-8, and Figure IV-9 show the proportion of students in each performance level in 2021 through 2023 by domain and grade. From 2022 to 2023, for listening, the level-4 percentages stayed the same in grades 1, 3, and 6; increased in kindergarten and grades 2 and 9; and decreased in the remaining grades. For speaking from 2022 to 2023, the level-4 percentages stayed the same in kindergarten and grade 1; increased slightly in grades 2, 3, 5–8, and 11; and decreased slightly in the other grades. For reading from 2022 to 2023, the level-4 percentages stayed the same in grade 1; increased in kindergarten and grades 2, 6, and 9; and decreased in the remaining grades. For writing from 2022 to 2023, the level-4 percentages stayed the same in kindergarten and grades 1, 3, and 9; increased in grades 2 and 6–8; and slightly decreased in the other grades.

Figure IV-6. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Listening

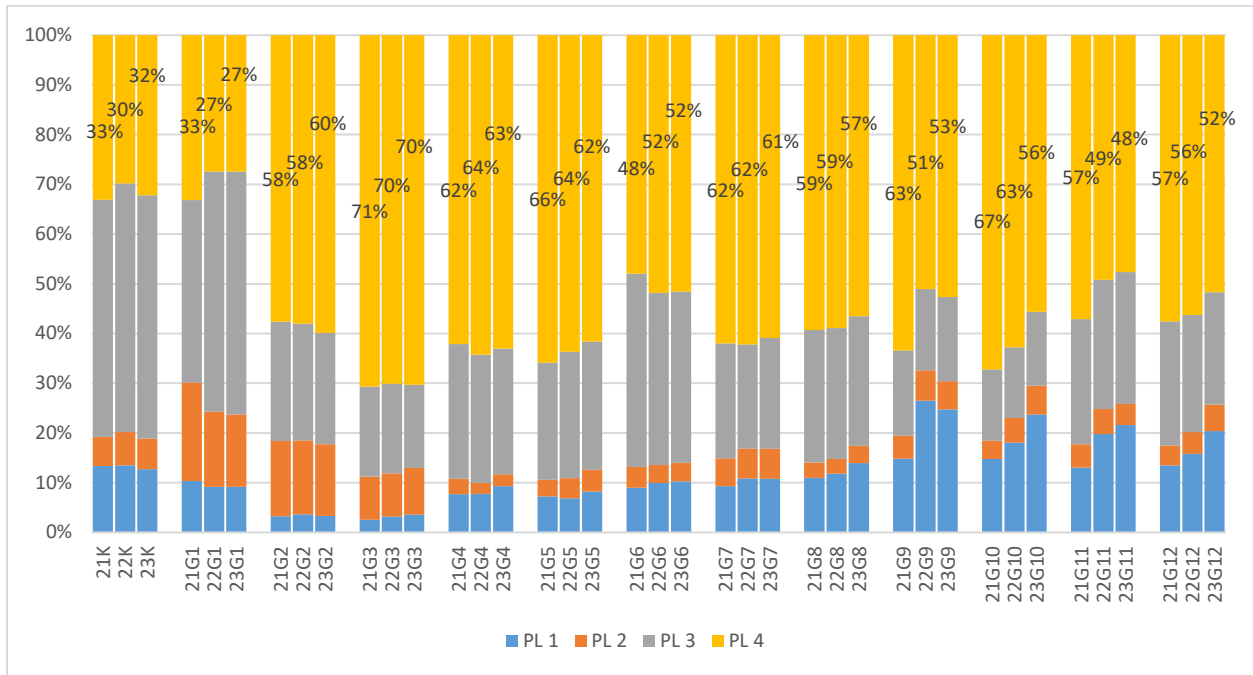


Figure IV-7. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Speaking

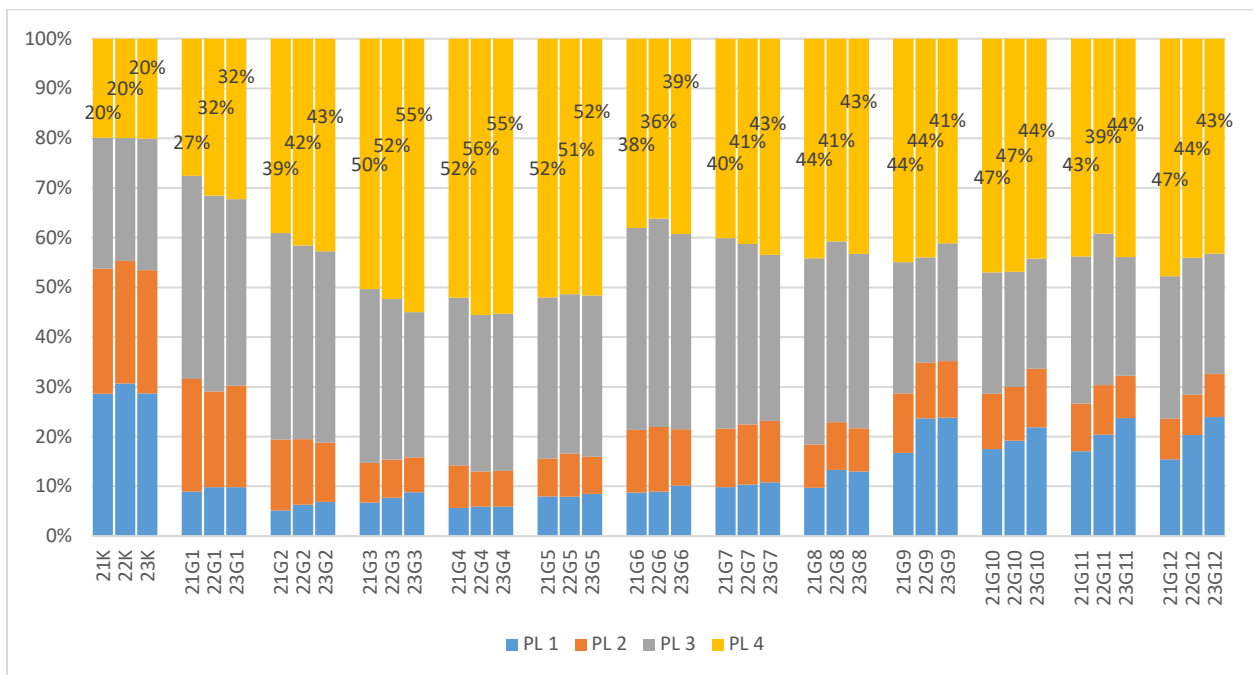


Figure IV-8. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Reading

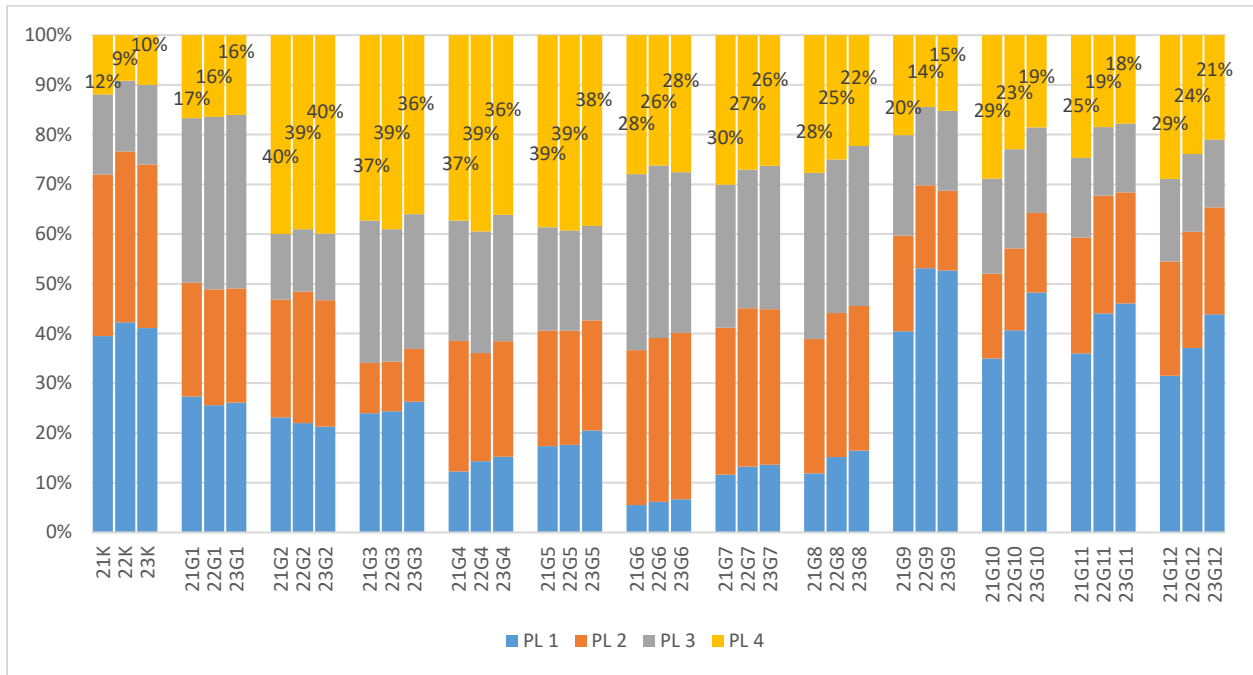
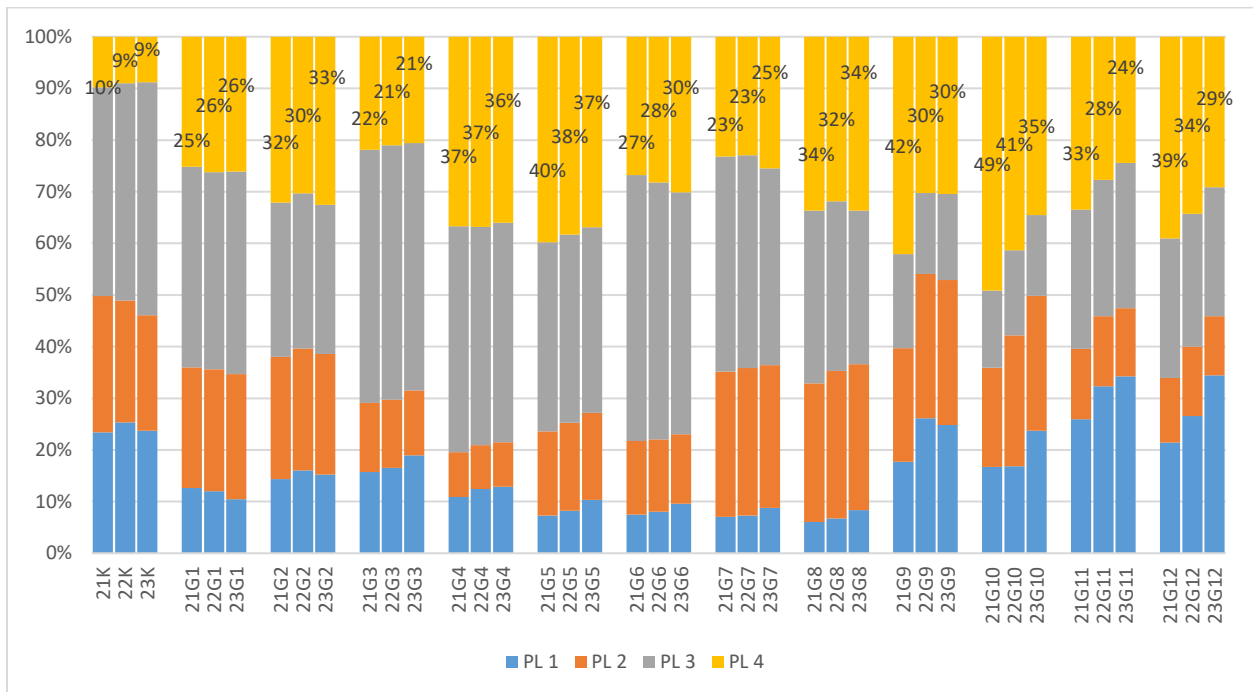
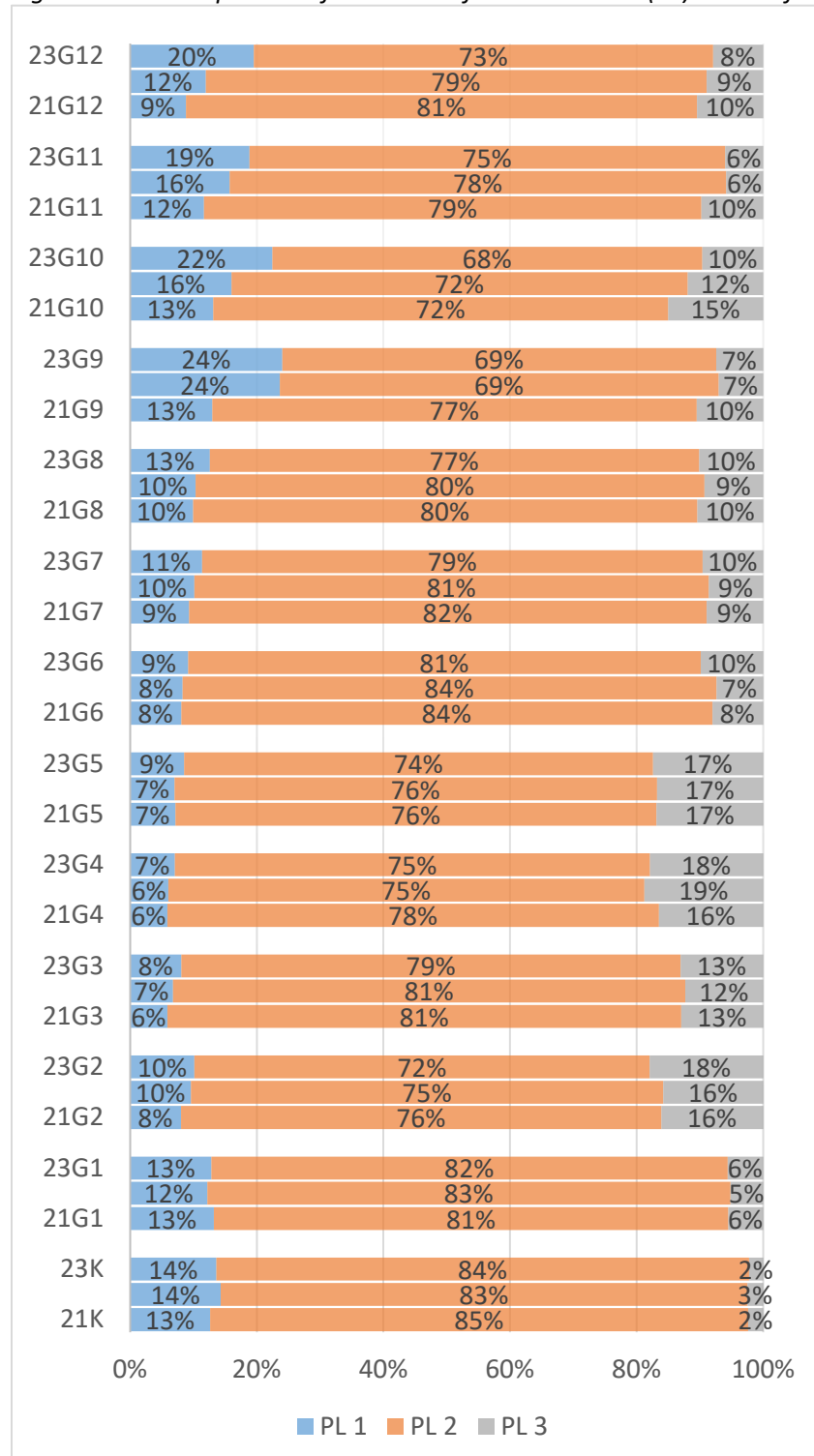


Figure IV-9. Comparison of Performance-Level (PL) Results from 2021 Through 2023 for Writing



The trend of the overall proficiency rates is provided in Figure IV-10. From 2022 to 2023, the overall proficiency rates stayed the same for grades 5, 9 and 11; increased slightly in grades 1–3 and 6–8, by 1% to 3%; and decreased in the other grades by 1% or 2%. The proficiency rates in grade 5 stayed the same for all three years from 2021 to 2023.

Figure IV-10. Comparison of Overall Performance-Level (PL) Results from 2021 Through 2023



## IV.3 Full Performance Continuum

The overall performance level of KELPA is a summary of students' performance in the four domains of listening, speaking, reading, and writing. The overall performance encompasses a full spectrum of student performance profiles in each of the four domains. There are 256 possible profiles for all domain performance, because each domain performance can be 1, 2, 3, or 4. For proficiency level 1, there are 16 possible profiles because each domain can be either 1 or 2. There is only one possible profile (4444) for proficiency level 3. Therefore, there are 239 ( $256 - 16 - 1$ ) possible profiles for level 2. Table IV-21 shows the number of performance profiles observed for overall proficiency by grade or grade band for proficiency levels 1 and 2. It shows that 16 profiles were observed for all grades and grade bands, except for grades 4–5 for students in proficiency level 1. The number of profiles observed ranges from 163 to 203 for students in proficiency level 2. The high number of profiles observed compared to all possible profiles indicated that KELPA results cover a full spectrum of student performance in the four domains.

*Table IV-21. Number of Performance Profiles by Grade or Grade Band*

Proficiency level	Grade or grade band	Number of profiles observed
1	K	16
	1	16
	2–3	16
	4–5	15
	6–8	16
	9–12	16
2	K	184
	1	163
	2–3	165
	4–5	163
	6–8	165
	9–12	203

Table IV-22 shows the top-five most observed domain performance profiles of students in descending order (i.e., top individual listed first and top five listed last) in overall proficiency level 1 and level 2. The top five profiles for proficiency level 1 accounted for 61–92% of students in that proficiency level across grade or grade band. The top five profiles for proficiency level 2 accounted for 18–35% of students in that proficiency level across grade or grade band. Across all grades, most overall level-1 students have a profile of 1111, corresponding to level 1 in listening, speaking, reading, and writing. The most observed profile for overall level-2 proficient students in grade K is 3111, indicating listening skills at level 3 and speaking, reading, and writing skills at level 1. For students in grades 2–5, most overall level-2 students have listening, speaking, and reading skills at level 4 and writing skills at level 3 (profile of 4443). For middle and high school grades, most level-2 students have listening, speaking, and writing skills at level 4 and reading skills at level 3 (profile of 4434).

Table IV-22. Domain Performance Profiles by Grade or Grade Band

Overall proficiency	Domain performance profiles in listening, speaking, reading, and writing					
	Grade K	Grade 1	Grades 2–3	Grades 4–5	Grades 6–8	Grades 9–12
Level 1	1111	1111	2111	1111	1111	1111
	2111	2212	1111	1211	1121	1211
	1112	2211	2211	2211	1211	1112
	1211	1211	2212	1112	1112	1212
	2112	1112	1211	1121	1222	2111
Level 2	3111	3333	4443	4443	4434	4434
	3323	3433	4433	4434	4433	4424
	4323	3323	4344	4433	4443	4443
	3223	4433	4343	4344	4333	4423
	4433	4434	4333	4343	3323	4433



## V. Inclusion of All Students

This chapter provides a summary of the frequency of accommodations used in the 2023 Kansas English Language Proficiency Assessment (KELPA) administration, as well as information about domain exemption in KELPA administration. For more detailed information about the accessibility framework in Kansas assessments, accessibility supports, available accommodations on KELPA, and the guidelines and procedures for selecting accommodations on KELPA, refer to [sections V.1 through V.3](#) in the *2020 KELPA Technical Manual* (Achievement and Assessment Institute [AAI], 2021a).

### V.1 Accommodations

All students who are identified as English learners, including those who need accommodations, must take KELPA. A three-tiered accessibility framework (i.e., Tier 1: Universal features for all students, Tier 2: Designated features for some students, Tier 3: Accommodations) is applied in Kansas state assessments; refer to [The Kansas Accessibility Manual](#). Accessibility tools, which vary by testing program, are available for all students taking various components of the Kansas assessments in the Kansas Assessment Program<sup>7</sup> (KAP). Without altering the assessment's validity, score interpretation, reliability, or security, assessment accommodations provide equitable access during assessments for students with disabilities. If the accommodation requested for a student changes the construct being tested, the test will not be valid for the student. Refer to [Section V.4.1 Selection of Accommodations](#) in the *2020 KELPA Technical Manual* (AAI, 2021a) for guidelines that are applied to every available accommodation on KELPA.

More details about KELPA accommodations can be found in the [KELPA Examiner's Manual](#), including an overview, prohibited practices, and recording accommodations used during testing (i.e., most testing accommodations should be entered into the student's Personal Needs Profile [PNP]). The [Kite Educator Portal Manual for Test Coordinators](#) provides additional information about accommodations for Kite® tools.

#### V.1.1 Selection of Accommodations

Individualized education programs (IEPs), 504 plans, services for English for speakers of other languages, and Student Improvement Team plans may use only accommodations documented on those plans; refer to the [KELPA Examiner's Manual](#) for details. According to the [Kite Educator Portal Manual for Test Coordinators](#), accommodations must be recorded in a PNP or in Access Profile in Educator Portal. To use an accommodation not listed in [Tools and Accommodations for the Kansas Assessment Program](#), the examiner should contact the District Test Coordinator, who will send the request to the Kansas State Department of Education (KSDE). Refer to [Section V.4.1 Selection of Accommodations](#) in the *2020 KELPA Technical Manual* for guidelines that apply to accommodation selection.

#### V.1.2 Frequency of Accommodations

In addition to accommodations that are built-in features of the Kite system, test administrators provide some accommodations that are allowed locally for KELPA. Any nonstandard accommodation requests and approvals are handled by KSDE. Because features in Kite are activated according to students' needs, teachers are required to mark those needs in the PNP. The PNPs submitted by teachers determine the availability of test accommodations for individual students. Table V-1 presents the number of students

---

<sup>7</sup> The Kansas Assessment Program provides general education assessments (i.e., assessments on English language arts, mathematics, and science), alternate assessments, career and technical education assessments, and KELPA.

who took KELPA in Kansas in 2023 and had PNP accommodations. The summary in the table shows that accommodations were requested for no students in kindergarten, for one student in grade 1, for 11 students in grade band 2–3, for 41 students in grade band 4–5, for 48 students in grade band 6–8, and for 74 students in grade band 9–12. The most frequent accommodation (i.e., 105 students) was auditory calming, which provides relaxing, peaceful background music while a student takes the test. The second- and third-most frequent accommodations (i.e., 35 and 14 students) were whole screen magnification and color contrast, respectively.

*Table V-1. Number of Students Using Accommodations by Grade or Grade Band*

Grade or grade band	No. of students using accommodation								Total
	ASL	Auditory calming	Color contrast	Color overlay	Masking	Reverse contrast	Switches	WSM	
K	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	1
2–3	0	7	0	0	0	0	0	4	11
4–5	5	18	2	1	0	0	4	11	41
6–8	0	34	4	2	0	1	0	7	48
9–12	1	46	8	1	1	1	4	12	74
Total	6	105	14	4	1	2	8	35	175

*Note.* ASL = American Sign Language, WSM = whole screen magnification.

## V.2 Domain Exemptions

In some situations, students may be exempt from taking a domain test. Special-circumstances codes available in Educator Portal allow school districts to manage test exemptions. Domain exemption requests were reviewed and approved by KSDE. Exempted domains were not included in the determination of overall proficiency. For example, students who are deaf or hard of hearing may be exempted from the listening test. For these students, overall proficiency will be determined by speaking, reading, and writing domain performance, and students will be considered proficient overall if they score at level 4 in the speaking, reading, and writing domains. Table V-2 shows the number of students exempted from testing by domain for the 2023 administration. Speaking is the most likely domain to be exempted from testing, with a total student count of 15 across all grades. Reading and writing are the least likely domains to be exempted from testing, with a total student count of no more than 1.

*Table V-2. Number of Students Exempted for Testing by Domain and Grade*

Grade	Listening	Speaking	Reading	Writing
K	0	5	0	1
1	0	1	0	0
2	0	2	0	0
3	0	3	0	0
4	0	4	0	0
5	0	2	0	0
6	1	0	0	0
7	1	0	0	0
8	1	0	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	1	0	0
Total	3	18	0	1

## VI. Academic Achievement Standards and Reporting

The Kansas English Language Proficiency Assessment (KELPA) standard-setting event occurred virtually in October 2020. The standard-setting event was composed of two major activities: the panelist advance training and assignments, and the virtual panel meetings to set cut scores. The Bookmark standard-setting method (Cizek & Bunch, 2007) was used to establish cut scores. For detailed procedures of the KELPA standard-setting event, as well as information about evaluations of the standard-setting method and event, refer to [Chapter VI](#) of the *2020 KELPA Technical Manual* (Achievement and Assessment Institute [AAI], 2021a). Because there were no updates to anything related to standard setting or performance level during the 2022–2023 school year, this chapter briefly updates information about student score reports.

### VI.1 Reporting

The 2023 KELPA testing window ended on March 10, 2023, and the scoring window closed on March 31, 2023. KELPA student reports were made available to all school districts on April 20, 2023, and in the Parent Portal on April 27, 2023.

#### VI.1.1 Student Reports

Performance levels for listening, speaking, reading, and writing were used to determine overall proficiency level, which is defined by the Kansas State Department of Education (KSDE). To be considered proficient (i.e., level 3 on overall proficiency) and eligible to exit the English for speakers of other languages (ESOL) program, students must receive 4s on all domain scores. Students who receive all 1s or 2s on the domain scores are considered not proficient (i.e., level 1 on overall proficiency). Students who do not meet the criteria for either level 1 or level 3 on overall proficiency are considered nearly proficient (i.e., level 2). In response to the COVID-19 pandemic and in consultation with KSDE and the Kansas Technical Advisory Committee, the following text was added to the top of the student report for both 2022 and 2023 administrations:

When interpreting student progress toward proficiency on the KELPA, please take into consideration how the conditions for learning, which may have been disrupted by the pandemic, may influence performance.

The 2023 KELPA student report kept the same format and information used in the 2022 student report. Both the overall proficiency level and the domain performance levels are provided in the student report. The overall proficiency levels are derived from student performance in the four domains.

### VI.1.2 Interpretive Guides

Descriptions of what students should know and be able to do at each performance level are provided in the reports. Nontechnical language is used to assist readers in interpreting the information in the reports. In addition, the [Educator Guide to KELPA Student Score Reports](#) and the [Parent Guide to KELPA Student Score Reports](#) (and its [Spanish translation](#)) are provided to assist the interpretation of the score reports. These guides explain the scores presented in the report and how the overall proficiency level and domain performance levels are determined. They also help readers understand students' progress toward English proficiency.

## VII. Ongoing Maintenance for KELPA Program

This chapter summarizes the ongoing program improvements and maintenance for the Kansas English Language Proficiency Assessment (KELPA).

### VII.1 Updates for the 2023 Administration

The multi-year enhancement effort for KELPA rater-training materials was completed during the 2022–2023 administration. In previous administrations, rater-training materials were only available for selected educator-scored items in speaking and writing. For the 2023 KELPA administration, rater-training materials were available for all educator-scored items in speaking and writing. The purpose of the updated materials is to support educators in applying rubrics to specific prompts, which enhances the validity of the constructed-responses item scores. For detailed information, refer to Section II.2.2 Development of Rater-Training Materials of the current manual.

### VII.2. Plans for Future Administration

#### VII.2.1 Multiple Test Forms

The KELPA program utilizes a pre-equated design where operational items with known item statistics are used to develop scoring tables prior to test administration. KSDE plans to use the same methodology to expand the KELPA item pool in the future. Newly developed items will be embedded in the operational test administration for field testing. The operational items will serve as linking items to place the field-test items on the KELPA item response theory (IRT) scale. Items in the expanded item pool will be used to develop new test forms for KELPA.

#### VII.2.2 Improve Reliability

Classification consistency and accuracy analyses at domain performance cut points provided information about whether KELPA provides accurate and reliable classification around the three performance cut points. The results of classification consistency and accuracy analyses may be used as guidance on item needs to improve classification accuracy on the level-3 and level-4 cut (proficient cut) when KSDE expands the KELPA items pool for future administrations. For example, the grade K listening test has a classification consistency of .77 between performance levels 3 and 4 (lower than the classification consistency between levels 1 and 2, and between levels 2 and 3), which indicated the current grade K listening test may need more difficult items to differentiate students at higher ability levels. Similarly, the grade K reading test has a lower classification consistency between performance levels 1 and 2, which indicates the need for easier items to differentiate students at lower ability levels.

#### VII.2.3 Linguistic Process

The KELPA item pool was developed between 2019 and 2020. Through an external item-review process, Kansas educators reviewed items for item content and fairness, and also informally evaluated the intended linguistic processing complexity needed for responding to the test items.

In the future expansion of the KELPA item pool, a formal evaluation of linguistic processing complexity will be included as part of the external item-review process, so that educators who have experience working directly with English learners (ELs) can evaluate whether KELPA items elicit the intended

linguistic-processing complexity from those students. The educator-reviewers will use a rubric to determine the linguistic-processing complexity across the receptive and productive language-processing domains, including reading, listening, speaking, and writing. Reviewer feedback will be solicited and documented for each newly developed item, then analyzed by item-development specialists trained in EL item development.

Items will be accepted when the response process is aligned with the intended linguistic process. Some items may be revised to better align with the intended linguistic process. In rare instances, items that require students to understand grammatical terms rather than apply knowledge will be rejected if they cannot effectively be revised to focus on the intended linguistic process.

## References

- Achievement and Assessment Institute. (2021a). *2020 KELPA technical manual*. Kansas State Department of Education. [https://ksassessments.org/sites/default/files/documents/technical-manuals/KELPA\\_Technical\\_Manual\\_2020.pdf](https://ksassessments.org/sites/default/files/documents/technical-manuals/KELPA_Technical_Manual_2020.pdf)
- Achievement and Assessment Institute. (2021b). *2021 KELPA technical manual*. Kansas State Department of Education. [https://ksassessments.org/sites/default/files/documents/technical-manuals/KELPA\\_Technical\\_Manual\\_2021.pdf](https://ksassessments.org/sites/default/files/documents/technical-manuals/KELPA_Technical_Manual_2021.pdf)
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. SAGE Publications.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>
- Elementary and Secondary Education Act, 20 U.S.C. § 6301 (1965). <https://govinfo.gov/content/pkg/COMPS-748/uslm/COMPS-748.xml>
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). <https://www.congress.gov/bill/114th-congress/senate-bill/1177>
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613–619. <https://doi.org/10.1177/001316447303300309>
- Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings* (ED532068). ERIC; Westat, Center for Educator Compensation Reform. <https://files.eric.ed.gov/fulltext/ED532068.pdf>
- 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. <https://www.jstor.org/stable/1435147>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.
- Sinclair, A. L., Paulsen, J., & Thacker, A. (2021). *Kansas English Language Proficiency Assessment: Alignment study*. The Human Resources Research Organization.



# Appendix A: 2023 KELPA Teacher Survey

## Introduction

This is a voluntary survey about the Kansas English Language Proficiency Assessment (KELPA) developed by the Achievement and Assessment Institute at the University of Kansas. You may stop participating at any time without penalty. You must be 18 or older to participate in this survey.

This survey's purpose is to provide feedback on teachers' and test administrators' testing experience with KELPA and on user experience with the technology. All responses are confidential, and results will be reported only to groups of respondents. No discomfort or risks to you are anticipated. No direct benefits for you are anticipated, though responses to this survey may be used to inform improvements to KELPA that may benefit students and teachers in the future.

This survey takes about 10 to 15 minutes to complete. You may withdraw from participating in the survey at any time. Please contact us via [kite-support@ku.edu](mailto:kite-support@ku.edu) if you have questions about your participation in the survey.

Your completion of the survey signifies your consent to participate. Thank you in advance for your participation. Your responses are valuable in helping improve the program.

## I. Demographics

1. Although you may serve many roles in your district, please select the one role that best describes your position as it relates to the Kansas English Language Proficiency Assessment (KELPA). [Single select]

- Building Test Coordinator (BTC)
- Building User (BU)
- curriculum director, curriculum coordinator
- district or building administrator
- District Test Coordinator (DTC)
- District User (DU)
- English learner (EL) instructional coach
- Program director or program coordinator
- Support staff
- Teacher (i.e., Classroom, Title 1, Special Education, EL) who administered KELPA
- Teacher (i.e., Classroom, Title 1, Special Education, EL) who did NOT administer KELPA
- Technology director or technology coordinator

2. If your role in KELPA is test administrator, for which grade or grade band did you administer KELPA this year? Please select all that apply.

- kindergarten
- grade 1
- grade band 2–3

- grade band 4–5
- grade band 6–8
- grade band 9–12

3. Please indicate the number of years of K–12 educational experience you have in each of the following areas. [limit input up to two-digit numbers only]

English language arts \_\_\_\_\_

Mathematics \_\_\_\_\_

Science \_\_\_\_\_

English learners \_\_\_\_\_

## II. Technology

The following questions are about the use of Kite® Suite applications to KELPA summative assessments in 2022–2023 administration.

**Kite Educator Portal (EP)** is used to manage data and score some speaking and writing items for KELPA summative assessments.

1. How is your experience in identifying students who were ready to take KELPA? Was it easy to see who was enrolled, who was rostered, and who was assigned tests? [OPEN ENDED]
2. How is your experience with KELPA scoring? Which aspects of the process were easy, and which were difficult? Please include your experience with second-rater scoring if it was used in your district. [OPEN ENDED]
3. Were you able to easily monitor testing progress and scoring progress? How is your experience with dashboards and/or data extracts that help monitor these statuses? [OPEN ENDED]
4. How is your experience in getting required test administration documents in the HELP tab? What was your experience uploading student writing samples through the SURVEY tab in EP? [OPEN ENDED]
5. We are always working hard to enhance the Kite Educator Portal. Are there qualities you would change or features you would like to add regarding KELPA? Please explain. [OPEN ENDED]

**Kite Student Portal (SP)** is used to deliver KELPA summative assessments to students.

6. Select the device type used by most assessed students within each grade or grade band.  
[single select]

	PC	Mac	Chromebook	iPad
Kindergarten				
Grade 1				
Grade 2–3				
Grade 4–5				
Grade 6–8				
Grade 9–12				

7. How is your experience in installing SP on district devices this year? [OPEN ENDED]
8. What is your overall opinion of students' use of SP to take KELPA summative assessments? Did you experience any issues with audio playback or recording student audio? If so, please explain. [OPEN ENDED]
9. Please provide other feedback related to your experience using SP for KELPA summative assessments. [OPEN ENDED]

### III. Rater Training for Scoring KELPA Items

10. Please rate the following statements about KELPA rater-training materials provided in your school district.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The rater-training materials helped me apply rubrics for scoring students' responses to speaking items.					
The rater-training materials helped me apply rubrics for scoring students' responses to writing items.					
The length of the state scoring window (human-scoring completed by 3/31/2023) was sufficient.					

11. Please rate the following statements about KELPA rater-training workshops provided in your school district.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The local rater training helped me understand the scoring rubrics.					
The local rater training helped me know how to use the scoring rubrics.					
The local rater training provided useful information for my role as a rater.					

The local rater training was well organized.					
The KSDE-published rater-training materials were easy to use.					
The KSDE-published rater-training materials helped me score responses confidently.					
The amount of time used for local rater training was about right.					

#### IV. Test-Administration Experience

1. Please rate the following statements about test administration for the **listening** domain.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The domain test length was appropriate for intended grade levels.					
The test instructions were clear.					
The test instructions were helpful to students.					

2. Please rate the following statements about test administration for the **speaking** domain.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The domain test length was appropriate for intended grade levels.					
The test instructions were clear.					
The test instructions were helpful to students.					



3. Please rate the following statements about test administration for the **reading** domain.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The domain test length was appropriate for intended grade levels.					
The test instructions were clear.					
The test instructions were helpful to students.					

4. Please rate the following statements about test administration for the **writing** domain.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The domain test length was appropriate for corresponding grade levels.					
The test instructions were clear.					
The test instructions were helpful to students.					

5. Please rate the following statements about your test-administration experience in general.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
I was confident in my ability to administer KELPA.					
The required test-administrator training prepared me for the responsibilities of a test administrator.					
The District Test Coordinator or Building Test Coordinator training sessions provided across the state were helpful.					

6. Please provide any suggestions to help us improve your ability to administer KELPA.  
[OPEN ENDED]

## V. Student Experience

Please rate the following statements about student experience.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The content of KELPA measured important English language proficiency knowledge, skills, and abilities.					
My student(s) had access to all necessary accessibility supports to participate in the assessment.					
In general, English learners (ELs) classified as <b>Proficient</b> according to their KELPA scores <b>can</b> fully access grade-level academic content.					
In general, ELs classified as <b>Not Proficient</b> according to their KELPA scores are <b>not</b> able to fully access grade-level academic content without English for Speakers of Other Languages (ESOL) services.					



## VI. Resources

Please rate the following statements about the KELPA support materials.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The <i>KELPA Examiner's Manual</i> was helpful.					
The <i>KELPA Scoring Manual</i> was helpful.					
The <i>KAP Practice Test Guide for Educators</i> was useful and helpful.					
The <i>KELPA Test Administration and Scoring Directions for Speaking</i> documents were helpful.					
The <i>KELPA Test Administration and Scoring Directions for Writing</i> documents were helpful.					

## VII. 2023 KELPA Update

An update to the summative KELPA program this year is the addition of school and district reports, in which district administrators and school administrators can view the performance of the entire district or school. If you are an administrator with access to these new reports, please respond to the following questions. Otherwise, you may opt out of this section.

1. Please rate the following statements about this update to KELPA.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
Overall, the KELPA school and district reports were helpful.					
Overall, the KELPA school and district reports were easy to understand.					
On the KELPA district report, displaying results from each school within the district was helpful.					

2. What other feedback do you have about the new KELPA school and district reports?  
[OPEN ENDED]

## Appendix B: Summary Results of Teachers' Responses to Survey Questions<sup>8 9</sup>

*Table B-1. Responses About Teachers' Role Relating to KELPA (N = 101)*

Role	N	%
Building Test Coordinator (BTC)	25	25
Building User (BU)	9	9
Curriculum Director, Curriculum Coordinator	0	0
District or Building Administrator	1	1
District Test Coordinator (DTC)	6	6
District User (DU)	2	2
English learner (EL) Instructional Coach	5	5
Program Director or Program Coordinator	3	3
Support staff	0	0
Teacher (i.e., Classroom, Title 1, Special Education, EL) who administered KELPA	50	50
Teacher who did NOT administer KELPA	0	0
Technology director or technology coordinator	0	0

*Table B-2. Distribution of Test Administrators by Grade or Grade Band (N = 101)*

Grade or grade band	%
Kindergarten	20
1	19
2–3	21
4–5	20
6–8	13
9–12	7

*Table B-3. Educators' Professional Experience in Years (N = 101)*

Years	Experience with (%)			
	English language arts	Mathematics	Science	English learners
0–2	23	29	38	10
3–5	2	6	9	9
6–9	6	6	6	15
10 or more	69	59	48	66

<sup>8</sup> Percentages in the tables may not sum to 100% because of rounding.

<sup>9</sup> Blank and “not applicable” responses were excluded from sample count N.

*Table B-4. Educators' Responses About Device Type used by Students in 2022–2023 (N = 101)*

	Device type (%)				
	PC	Mac	Chromebook	iPad	No Response
Kindergarten	4	2	32	35	28
Grade 1	4	2	33	33	29
Grades 2–3	4	2	48	21	26
Grades 4–5	4	2	57	11	26
Grades 6–8	3	4	40	10	44
Grades 9–12	3	3	31	2	61

*Table B-5. Educators' Responses About KELPA Rater-Training Materials (N = 101)*

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
The rater-training materials helped me apply rubrics for scoring students' responses to speaking items.	100	1	1	17	75	6
The rater-training materials helped me apply rubrics for scoring students' responses to writing items.	101	1	0	15	76	8
The length of the state scoring window (human-scoring completed by 3/31/2023) was sufficient.	98	1	2	5	90	2

Table B-6. Educators' Responses About KELPA Rater-Training Workshops (N = 101)

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
The local rater training helped me understand the scoring rubrics.	101	1	3	10	62	24
The local rater training helped me know how to use the scoring rubrics.	101	1	3	6	66	24
The local rater training provided useful information for my role as a rater.	100	2	1	6	67	24
The local rater training was well organized.	99	0	1	7	68	24
The KSDE-published rater-training materials were easy to use.	101	1	4	17	67	11
The KSDE-published rater-training materials helped me score responses confidently.	101	1	2	19	68	10
The amount of time used for local rater training was about right.	100	1	1	12	65	21

Table B-7. Educators' Responses About KELPA Test-Administration Experience for Each Domain (N = 101)

	Domain	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
The domain test length was appropriate for intended grade levels.	Listening	99	1	8	17	74	0
	Speaking	99	1	7	7	85	0
	Reading	98	1	7	20	71	0
	Writing	99	0	6	13	80	1
The test instructions were clear.	Listening	98	1	1	11	87	0
	Speaking	99	1	2	18	79	0
	Reading	98	1	2	12	85	0
	Writing	99	1	2	15	81	1
The test instructions were helpful to students.	Listening	99	2	5	26	67	0
	Speaking	99	1	6	23	70	0
	Reading	97	1	4	26	69	0
	Writing	98	1	4	24	69	1

Table B-8. Educators' Responses About KELPA Test-Administration Experience in General (N = 101)

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
I was confident in my ability to administer KELPA.	100	0	0	3	97	0
The required test-administrator training prepared me for the responsibilities of a test administrator.	100	0	1	6	91	2
The District Test Coordinator or Building Test Coordinator training sessions provided across the state were helpful.	99	0	1	8	69	22

Table B-9. Educators' Responses About Student Experience (N = 101)

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
The content of KELPA measured important English language proficiency knowledge, skills, and abilities.	98	3	5	44	47	1
My student(s) had access to all necessary accessibility supports to participate in the assessment.	99	3	2	10	84	1
In general, English learners (ELs) classified as <b>Proficient</b> according to their KELPA scores <b>can</b> fully access grade-level academic content.	98	6	2	26	65	1
In general, ELs classified as <b>Not Proficient</b> according to their KELPA scores are <b>not</b> able to fully access grade-level academic content without English for Speakers of Other Languages (ESOL) services.	98	3	14	28	53	2

Table B-10. Educators' Responses About KELPA Support Materials (N = 101)

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
The <i>KELPA Examiner's Manual</i> was helpful.	100	0	2	10	88	0
The <i>KELPA Scoring Manual</i> was helpful.	100	1	1	12	85	1
The <i>KAP Practice Test Guide for Educators</i> was useful and helpful.	100	0	2	10	77	11
The <i>KELPA Test Administration and Scoring Directions for Speaking</i> documents were helpful.	98	1	1	16	82	0
The <i>KELPA Test Administration and Scoring Directions for Writing</i> documents were helpful.	99	1	1	16	82	0

Table B-11. Educators' Responses About KELPA Update (N = 101)

	N	Disagree (%)	Somewhat Disagree (%)	Somewhat Agree (%)	Agree (%)	Not Applicable (%)
Overall, the KELPA school and district reports were helpful.	86	5	6	8	59	22
Overall, the KELPA school and district reports were easy to understand.	85	2	2	16	58	21
On the KELPA district report, displaying results from each school within the district was helpful.	83	2	2	12	54	29

## Appendix C: Responses to Open-Ended Summative Educator Survey Questions

In the 2023 Kansas English Language Proficiency Assessment (KELPA) summative educator survey, some open-ended questions were asked to collect educators' feedback and/or opinions on topics, such as the usability of Kite® Suite applications to KELPA summative assessments in 2022–2023 administration and the usage of the new KELPA school and district reports. Below are a list of these questions and a summary of responses to them.

**Q10. *How is your experience in identifying students who were ready to take KELPA? Was it easy to see who was enrolled, who was rostered, and who was assigned tests?***

- Most educators respond that it was easy to accomplish this task.
- Educators indicate that with assistance of test coordinator, building administrator, school district, and PowerSchool, this task was easy.
- Educators would like to get direct access to students' KELPA status on the Kite Educator Portal.

"My issue has been when we have transfer students from another KS school, we do not always know who previously took and/or passed the KELPA. It does not immediately populate under the 'KELPA stud. scores current stud.,' so I test kids that are proficient."

"The difficulty is to make sure PowerSchool is correct. I wish there was a list of those who passed in the last 5 years to clean up and double check."

**Q11. *How is your experience with KELPA scoring? Which aspects of the process were easy, and which were difficult? Please include your experience with second-rater scoring if it was used in your district.***

- Educators complain about excessive scrolling when locating information, improper font size for window fitting, and too many clicks before submitting a score.
- Educators indicate that second-rating scoring was confusing and difficult.

"The amount of clicks you have to make for each question is too many. It seems you could do one question for all kids that you did simultaneously or deferred."

"The process is difficult to use because we have to continually scroll up and down on two different bars."

"Selecting the style of simultaneous each question is frustrating. The screens are not the correct size to see the questions along with the grading scale."

"It would be better if the recordings and examples were available at each question instead of having to find it in the PDF or in downloads."

"We feel that the 2nd rater is difficult to track down who needs what. There really should just be a 2nd rater tab that says exactly who needs it in each district. There's too much filtering school by school, grade by grade, etc."



**Q12. Were you able to easily monitor testing progress and scoring progress? How is your experience with dashboards and/or data extracts that help monitor these statuses?**

- Overall, educators appreciate the dashboards and data extracts for monitoring progress.
- Educators desire individual progress to be displayed.
- For educators in charge of multiple buildings, they would appreciate being able to run the district report all at once.
- Educators desire the dashboards to be updated sooner.
- Educators report that dashboards shrink test sessions on the Mac screener.
- Educators desire to see all students in a report for each domain, as well as their progress.
- Educators report that people who lack experience with KELPA test administration might find it difficult.

“It was easy to monitor student progress during testing sessions. Data extracts are confusing. It seems there are multiple reports that can be pulled for UNs and PWs. But, there is really only one helpful but it has the same students listed multiple times.”

“I see the writing and speaking progress, but teachers don't have access to see reading and listening progress at all. I don't know if something has submitted properly unless I go bother an administrator.”

“As a building test coordinator, I am only able to see the speaking and writing. I have to check with the building principal about the reading and listening to see if all students have completed the tests.”

“It was hard as a test administrator to monitor the testing progress of the screener, as I had to keep going to my district coordinator to get results. This limited me in knowing how I should proceed and delayed the process a little bit.”

**Q13. How is your experience in getting required test administration documents in the HELP tab? What was your experience uploading student writing samples through the SURVEY tab in EP?**

- The majority of educators report no issues with the HELP tab and find it easy to handle uploading.

“My experience with uploading scores through the survey tab was intimidating. It went smoothly once the upload was complete. The step-by-step directions are very thorough on how to maneuver through this process.”

**Q14. We are always working hard to enhance the Kite Educator Portal. Are there qualities you would change or features you would like to add regarding KELPA? Please explain.**

- Educators point out that scrolling down to locate answer choices was difficult to maneuver for young students, that clicking and dragging were difficult for them, and that the matching boxes were confusing.
- Educators report that having audio and questions on separate tabs was confusing.
- Educators desire less clicking and scrolling for each question when grading a test.
- Educators suggest listed testers to be able to monitor testing status.

- Educators suggest a summary screen to oversee the progress of all buildings under their charge.
- Educators indicate that a procedural guide or a step-by-step flowchart would be helpful.
- Educators desire the graphics to be more modern and appealing to children.
- Educators would like to have a default setting for deferred scoring on the speaking section to avoid repetitive clicking that could be reduced.
- Educators desire a security warning to be given to students when a locked browser is accessed.
- Educators suggest that students who need second rating should only show up under the second-rater tab.
- Educators would like to have to same access to download all the training materials and manuals as test coordinators do.
- Educators would like to have faster updates about student status and organization of test-day materials.

“The questions need to be visible on one screen. Students should not have to scroll up and down in order to see each answer. There were times when they couldn't see the question after scrolling down to see the answers.”

“There were a lot of bugs which made it take a long time to actually be able to administer the test. These issues should be worked out before testing. Since we start testing first (and it occurs during spring break) we don't have time to lose.”

**Q21. *How is your experience in installing SP on district devices this year?***

- Some educators report confusion in finding directions on the assessment.org website.
- Educators report that installing district-level software onto students' school-issued devices caused unresolvable error messages at a rate of about 10%.
- Many educators appreciate the assistance from their dedicated technology team, which handled the installation.

“This year was our smoothest on record.”

**Q22. *What is your overall opinion of students' use of SP to take KELPA summative assessments? Did you experience any issues with audio playback or recording student audio? If so, please explain.***

- Quite a few educators report difficulties in audio and recording responses.
- Educators report that switching devices, re-plugging headphones, re-login and resetting devices sometimes solved the recording issues.
- Educators report that dragging/dropping items and matching boxes could be difficult for lower grade students, especially on iPad.
- Educators report that using SP on desktop computers seemed much smoother than before.
- Educators report that iPads were prone to recording issues.
- Educators report that in-advance testing on microphones and headphones was essential for successful test administration.

“YES I did, a lot of red screens came up in the screener losing answers already recorded also had red screen in KELPA. The KELPA I had problems with going from the Listening test right into Reading ...no sound.”

"[I]t is hard to make sure younger kids are able to record without missing some of their speaking. Often kids somehow cut it off, which results in a low score when they shouldn't."

"There were several students who could not get the recording to work while taking the practice test. They were not able to practice because it would not record."

"During the speaking/writing portion students show the educator that the test session is complete--blue dots. However, during scoring it is evident that the test was not completed. Why are the blue dots shown when the test is not actually complete?"

"There were technical issues with sound on many of the students tests."

"I had my students use iPad for reading, listening, and writing. Due to so many recording issues last year on the iPad, I used desktops for speaking. I did not have near the issues of no audio files found this year as last."

"I wish they could figure out how to have everything on one screen and not have to scroll down. The scroll bar disappears and it is difficult for kids who aren't tech savvy."

"Students had to talk very loudly for it to record."

"We had trouble using our microphones this year so we had to use our internal microphone so that was interesting. So recordings were a little muffled because kids sat back away from computer."

"There was an issue with one of my students about recording her answers on the speaking. It uploaded her first answer, but not any of the others. I told the district test coordinator and he reset it and the student was able to do the speaking again."

"There were a few problems with the speaking responses not saving. We ended the students and then had them try the next day. Then, it worked."

"We tested all of our headphones with microphones prior to the assessment day, so we did not experience any issues with sound/audio not working."

"We had a lot of red screens the first day, then it was better. Audio was fine."

"I had experience with tapping sounds, the microphone cutting out, background noise and students speaking too quietly, which made it difficult to score on some occasions."

"The student portal could be easier to manipulate with touchscreen chromebooks."

"We had a few issues with iPads and it just spinning and then not saving their work."

"Yes sometimes the recording didn't record the student. It would be great if the students were warned if the recording wasn't saved."

"A few issues here and there - what if KITE could 'detect' if/when a student records a blank response and then alerted them to that fact on the end/review screen. That would be awesome."

**Q23. Please provide other feedback related to your experience using SP for KELPA summative assessments.**

- Educators would appreciate being able to view content in one page instead of scrolling.
- Educators would like shorter turnaround time to receive student results.
- Educators would like a quicker method to exit the student portal.

"It is harder for students when they have to move the screen up or down to read or answer the questions"

"On the ipad it would be nice to have the answer choices all in the same view vs the student having to scroll down to see all the answer choices. (Especially for primary K-2 grades)"

"I SO WISH that the entire picture and question bank could be seen on one page. Those chromebooks are so small that the kids are scrolling all over the place to see the picture then try to find the answer. We see no need to have an 'audio' page w/ no Qs."

"Kindergarten students can not drag and drop when the entire answer is not on the screen."

"Reiterating the possibility of receiving student results sooner - preferably before the school year is over."

"It is sometimes hard for students to answer when the entire question and answers cannot be seen altogether on one screen. It is sometimes difficult for younger ones to scroll up and down repeatedly to match answers or go back to refer to the question."

"There are a lot of steps to exit out of the Student Portal. It would be nice if there were a quicker method for exiting."

**Q62. What other feedback do you have about the new KELPA school and district reports?**

- Educators would appreciate communication with parents in multiple languages.
- Educators would appreciate earlier KELPA results.
- Educators express that prior KELPA records for transfer students would help prevent unnecessary testing on proficient students.
- Educators desire more informative or broken-down results to improve their future classroom instruction.
- Educators expect the language used by the KELPA report and the Kansas State Department of Education (KSDE) standards to be aligned.
- Educators desire screener results to be reported by domain as well.

"It would be nice if we could get scores earlier after the test window closes. There is too much time between the test and the results."

"Please provide KELPA results before the end of the school year. May1st would be even better!!"

"It would be helpful to receive the scores sooner than later. In addition, it would be helpful if the students passed and were proficient in an area, they wouldn't be required to take that portion of the test the following year."

"When students transfer to a new school in KS, I need to know if they previously took and/or passed the KELPA. This information is not readily available when a kid is new to our district, leading us to unnecessarily test proficient kids. Please fix this."

"I don't know that we receive enough information to help our students. A student scored a 2- does that mean they have very little English or they only answered 1 part of a 2 part question? It is hard to change instruction when the categories are so broad."

"I would like to receive detailed results of each student's scoring after the KELPA. I would love to know what my students missed to guide my instruction for them."

"Our district reports should align with the KELPA reports. We should use the same language for levels 1-5 based on KSDE standards and match KELPA levels to those standards."