



# KELPA

Kansas English Language Proficiency Assessment

## KELPA Technical Manual

October 2022

Copyright © 2022, 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.....	5
II.3.1	KELPA Teacher Survey.....	6
II.4	Test Security.....	6
III.	Technical Quality—Validity .....	7
III.1	Validity Evidence Based on Test Content.....	7
III.1.1	A Brief Summary of Alignment Activity Results .....	7
III.1.2	Postalignment Activities.....	8
III.2	Validity Evidence Based on Relations to Other Variables .....	8
III.3	Validity Evidence Based on Consequences of Testing .....	11
III.3.1	Teacher Survey.....	11
IV.	Technical Quality—Other.....	14
IV.1	Reliability .....	14
IV.1.1	Test Reliability.....	14
IV.1.2	Classification Consistency and Accuracy .....	16
IV.1.3	Interrater-Agreement Study .....	18
IV.2	Scoring and Scaling .....	23
IV.2.1	Operational Test Results.....	23
IV.2.2	Trend Data .....	36
IV.3	Ongoing Program Improvement.....	40
IV.3.1	Enhanced Rater-Training Materials Development .....	40
IV.3.2	Domain-Score Exemption .....	41

IV.3.3 Incident Response.....	41
V. Inclusion of All Students.....	42
V.1 Accommodations .....	42
V.1.1 Selection of Accommodations .....	42
V.1.2 Frequency of Accommodations .....	42
VI. Academic Achievement Standards and Reporting .....	44
VI.1 Reporting .....	44
VI.1.1 Student Reports .....	44
VI.1.2 Interpretive Guides .....	44
References .....	45

## Table of Tables

Table II-1. Updates to Rater-Training Materials .....	4
Table III-1. Correlations Between KELPA Domain Scores and KAP English Language Arts (ELA) Scores by Grade.....	9
Table III-2. Correlations Between KELPA Domain Scores and KAP Mathematics Scores by Grade .....	9
Table III-3. Correlations Between KELPA Domain Scores and KAP Science Scores by Grade .....	10
Table III-4. Performance of Proficient English Learners on KAP English Language Arts, Mathematics, and Science Assessments.....	11
Table IV-1. Coefficient Alpha by Domain and Grade or Grade Band .....	15
Table IV-2. Coefficient Alpha for Student Groups by Domain and Grade or Grade Band .....	16
Table IV-3. Classification Consistency (C) and Accuracy (A) by Domain and Grade .....	17
Table IV-4. Available Scoring Methods for Speaking and Writing .....	19
Table IV-5. Numbers of Districts, Schools, and Students Selected for Two Ratings.....	20
Table IV-6. Number of Students With Two Ratings by Domain and Grade or Grade Band.....	20
Table IV-7. Rater Agreement on Writing Items Scored Using the Individual Scoring Method by Grade or Grade Band .....	22
Table IV-8. Rater Agreement on Speaking Items .....	22
Table IV-9. Summary of Quadratic Kappa Classifications .....	23
Table IV-10. 2022 KELPA Participation Rates by State Board of Education (SBOE) District and Grade or Grade Band .....	25
Table IV-11. Percentage of Tested Students by Demographic Characteristic and Grade .....	26
Table IV-12. Scale-Score Descriptive Statistics by Grade for Listening.....	27
Table IV-13. Scale-Score Descriptive Statistics by Grade for Speaking.....	27
Table IV-14. Scale-Score Descriptive Statistics by Grade for Reading .....	28
Table IV-15. Scale-Score Descriptive Statistics by Grade for Writing .....	28
Table IV-16. Demographic Group Scale-Score Descriptive Statistics by Grade for Listening .....	33
Table IV-17. Demographic Group Scale-Score Descriptive Statistics by Grade for Speaking .....	34
Table IV-18. Demographic Group Scale-Score Descriptive Statistics by Grade for Reading.....	35
Table IV-19. Demographic Group Scale-Score Descriptive Statistics by Grade for Writing .....	36
Table IV-20. Number and Percentage of Enrolled and Tested Students by Grade: 2020 through 2022....	37
Table IV-21. Number of Students Exempted From Testing by Domain and Grade .....	41
Table V-1. Number of Students Using Accommodations by Grade or Grade Band .....	43

## Table of Figures

Figure IV-1. Performance-Level Results for Listening, 2022 Administration .....	29
Figure IV-2. Performance-Level Results for Speaking, 2022 Administration .....	29
Figure IV-3. Performance-Level Results for Reading, 2022 Administration .....	30
Figure IV-4. Performance-Level Results for Writing, 2022 Administration .....	30
Figure IV-5. Overall Performance-Level Results (2022 Administration) .....	31
Figure IV-6. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Listening .....	38
Figure IV-7. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Speaking .....	38
Figure IV-8. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Reading .....	39
Figure IV-9. Comparison of Performance-Level (PL) Results from 2020 Through 2022 for Writing .....	39
Figure IV-10. Comparison of Overall Proficiency-Level (PL) Results From 2020 Through 2022 .....	40

# 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 who are 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)—reading, speaking, listening, 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 2021 administration was the third administration of KELPA that was 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, as required by Title I of the Elementary and Secondary Education Act (ESEA; Every Student Succeeds Act, 2015). As part of the ESEA Title I accountability requirement, 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 language, and skills for using language that are 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 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](#) (AAI, 2021a). A detailed summary of accommodations is 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 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 2022 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](#) (AAI, 2021a).

## II. Assessment System Operations

This chapter provides updated information about KELPA design and development, administration, and test security. For more details (e.g., monitoring test administration), refer to Chapter II 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 Achievement and Assessment Institute (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](#) (AAI, 2021a), which also provides detailed information about KELPA test blueprints (i.e., [Section II.1.1 Test Blueprints](#)), test design (i.e., [Section II.1.2 Test Design](#)), and test construction (i.e., [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* (AAI, 2021a) 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 the rubric and 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* (AAI, 2021a) 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, 2021, and 2022 administrations.

## II.2.2 Development of Rater-Training Materials

This section describes the development of updated rater-training materials for the 2022 KELPA administration, as well as plans for the continuation of the staged roll-out to ensure all constructed-response (CR) items on the assessment include rater-training materials.

### II.2.2.1 Materials for the 2022 Administration

In 2021, the rater-training materials included one set of materials (anchor, calibration, and practice responses) for one writing prompt and one speaking prompt per grade or grade band. For the 2022 administration, these materials were expanded to include additional prompts, as noted in Table II-1.

*Table II-1. Updates to Rater-Training Materials*

Grade or grade band	No. of writing prompts	No. of speaking prompts
K	1	3
1	1	3
2–3	1	3
4–5	1	3
6–8	2	3
9–12	1	3

AAI content-development staff evaluated student responses from the 2020 test administrations, according to each rubric. Three sets of responses per item were selected for use in the materials that district and building coordinators use to train and calibrate local raters. Those sets consisted of an anchor set, a practice set, and a calibration set. The anchor set contains three responses for each score point (0–3) on the rubric to identify how the holistic rubrics are applied to a variety of student responses. There are 12 responses in the anchor set for the same item; each anchor-set response is accompanied by an explanation for the assigned score point. Each practice set and calibration set has 10 responses for the same item, with three responses at score points 3, 2, and 1 and one response at 0. Thus, for each item (or prompt) in each grade or grade band, there are 32 responses: 12 anchor-set responses, 10 practice-set responses, and 10 calibration-set responses. Both the calibration and practice sets are intended to help local raters practice; that is, they aid raters in developing an understanding of how to operationalize the rubrics by evaluating student-response examples at each score point.

In October 2021, the rater-training materials underwent an external review by Kansas educators and KSDE staff. There were three review panels (i.e., kindergarten and grade 1, grades 2–5, grades 6–12), and each panel examined both writing and speaking rater-training materials. Two educators served on each panel, along with KSDE staff members. The external reviews began with an orientation to familiarize the panelists with the task and their responsibilities. Following the orientation, panelists asynchronously reviewed all responses (i.e., the anchor and calibration sets) in their grade or grade band for both domains and sent their feedback to AAI’s content-development staff. Feedback included whether they agreed with the assigned score point or whether the score point was too low or too high, and whether they felt any revision was needed for the anchor-set explanations. Panels then met for

synchronous discussion of that feedback. Panelists discussed responses that they had rated differently from the rating given in the materials. When the panelists agreed that a response was not suitable for the assigned score point, AAI content-development staff showed (i.e., for writing) or played (i.e., for speaking) other preselected options based on scoring notes from the earlier process of response evaluation for the sets. Other options were presented until one of the new responses was determined by the panel to accurately demonstrate the knowledge and skills associated with the score point.

Using feedback in the synchronous discussions as well as asynchronous feedback on anchor-set explanations, AAI content-development staff made changes to the materials (mainly, replacing responses, revising explanations, reordering anchor-set responses) and finalized the documents for the 2021 KELPA administration.

Additionally, KSDE staff reviewed each of the responses for the practice and calibration sets. Like the process noted above, KSDE staff indicated whether they agreed with the assigned score point or whether the score point was too high or too low. If the response does not match the assigned score point, KSDE staff chose a student response for that score point from additional responses.

#### *II.2.2.1 Materials for the 2023 Administration*

The development process for 2023 materials will be similar to the process used for the 2022 materials. Content-development staff will select responses for all sets and write explanations for the anchor-set responses. During external reviews, educators will review anchor and calibration sets, and KSDE staff will review all sets. AAI content-development staff will use synchronous and asynchronous feedback to select any needed replacements, which KSDE will review.

## II.3 Test Administration and Scoring

The 2022 KELPA testing window was open to students from January 31 through March 11, 2022. Educators were able to enter scores for CR items until March 31, 2022. 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* (AAI, 2021a).

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. District coordinators 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* (AAI, 2021a). The training webinars, recorded and posted on the site, are provided, and updated every year. The training slides, frequently asked questions, and responses to these questions are also posted on the [DTC Virtual Training Webinars](#) site.

The standardized test-administration procedures provided for districts, schools, and teachers are described in the [2021–2022 KELPA Examiner’s Manual](#) (*Examiner’s Manual* hereafter). The *Examiner’s Manual* also provides guidance and procedures related to the administration of KELPA in 2021–2022, for example, 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 in [Section II.3.2 Test-Administration Procedures](#) of the *2020 KELPA Technical Manual* (AAI, 2021a).

### II.3.1 KELPA Teacher Survey

At the beginning of the KELPA testing window, KSDE sent out a notification of teacher surveys through KSDE email distribution lists to educators and encourage educators to participate in the KELPA teacher survey. 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, 2022. The survey (see Appendix A) included questions about educators' background and their experience with Kite, scoring, test administration, students' testing experience and supporting materials (e.g., the [2021–2022 KELPA Examiner's Manual](#), *KELPA Test Administration and Scoring Directions* for speaking and writing, etc.), learning and instruction in 2021–2022, and the utility of KELPA. One hundred forty-nine educators (7% of active Educator Portal users to whom students were rostered to take the 2022 KELPA) responded to the survey. Results of the teacher survey are included in Section III.3.1 Teacher Survey of the current manual.

## 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* (AAI, 2021a). 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* (AAI, 2021a) provide detailed information about and requirements for test-materials security, test-related data security, security of personally identifiable information, and accommodations-related security.

### III. Technical Quality—Validity

The 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), according to. There are five sources of evidence to consider when evaluating test-score validity (APA et al., 2014): evidence based on (a) test content, (b) response processes, (c) internal test structure, (d) relationships between test scores and other variables, and (e) consequences of testing. The test forms in 2022 were the same as the operational forms in 2021 and 2020; 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 2021–2022 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). The independent study collected information to address six claims:

1. KELPA items are aligned to 2018 Standards.
2. KELPA items represent the 2018 Standards.
3. KELPA meets test blueprints, representing a balanced assessment.
4. KELPA domain-level tests are reliable <sup>1</sup>.
5. KELPA includes items representing a range of linguistic difficulty levels.
6. Language proficiency requirements of the academic standards are addressed by the 2018 Standards.

##### III.1.1 A Brief Summary of Alignment Activity Results

Results of the alignment study indicate that (a) the criterion applied to Claim 1 was met; (b) the criterion applied to Claim 2 was met for most grades or grade bands; (c) the criterion applied to Claim 3 was met in all four domains for grade bands 6–8, for all domains except reading in grade band 9–12, and two of the four domains in grade 1 and grade bands 2–3 and 4–5; (d) the criterion applied to Claim 5 was met in all or most domains for kindergarten and grade bands 2–3 and 4–5, but not for grade 1 and other grade bands; and (e) the criterion applied to Claim 6 was met for all grades or grade bands and academic content areas except grade-1 mathematics. For more details, refer to [Section III.1 Validity Evidence Based on Test Content](#) in the *2021 KELPA Technical Manual* (AAI, 2021b). For full results of the

---

<sup>1</sup> Claim 4 of the alignment study is addressed by domain-level test reliabilities reported in the [2020 KELPA Technical Manual](#).

alignment study, refer to the *Kansas English Language Proficiency Assessment: Alignment Study* (Sinclair et al., 2021).

### III.1.2 Post alignment Activities

Based on the findings of the KELPA alignment study, Achievement and Assessment Institute (AAI) content-development staff will review the partially aligned items and other items to update item alignment. After the review, the AAI psychometric team will analyze cluster-level coverage for the target items. The two teams will then consider the results from the cluster-level coverage and discuss whether they would merit blueprint updates. If blueprint updates are merited, the recommendations will be sent to the Kansas State Department of Education (KSDE) for review and approval. If no blueprint updates are needed, AAI psychometric and content-development teams will collaboratively report the findings from the post alignment activities and submit the report as a peer-review response memo.

## 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 between students’ performance on different assessments measuring similar constructs. *Discriminant evidence* is provided by relationships between 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-1 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 .65 (grade 4); the weakest correlations were observed between ELA and the speaking domain, ranging from .21 (grade 10) to .30 (grade 6). Correlation coefficients between KAP ELA and KELPA speaking domain across grades were small (except in grade 6). For relationships between KAP ELA and KELPA listening, reading, and writing, medium to large correlation coefficients were found across grades.

*Table III-1. 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	.46	.30	.61	.55
4	.54	.26	.65	.58
5	.46	.24	.58	.52
6	.52	.30	.63	.49
7	.52	.26	.59	.44
8	.45	.24	.51	.40
10	.38	.21	.53	.45

Table III-2 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 .21 (grade 10) to .55 (grade 3); the weakest correlation was between KAP mathematics and KELPA speaking domain, ranging from .07 (grade 10) to .30 (grade 3). Relationships between KAP mathematics and KELPA were weakest for grade 10.

*Table III-2. Correlations Between KELPA Domain Scores and KAP Mathematics Scores by Grade*

Grade	Correlation between KAP mathematics and domain			
	Listening	Speaking	Reading	Writing
3	.46	.30	.55	.56
4	.46	.21	.50	.49
5	.42	.19	.43	.42
6	.39	.20	.47	.34
7	.39	.20	.42	.34
8	.33	.23	.37	.33
10	.17	.07	.21	.22

Table III-3 presents correlations between KELPA domain scores and KAP science scores. The strongest correlation was between KAP science and reading scores, ranging from .35 (grade 11) to .50 (grade 5); the weakest correlation was between science and speaking scores, ranging from .13 (grade 11) to .26 (grade 5). Correlations between KAP science and KELPA scores is weakest for grade 11.

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

Grade	Correlation between KAP science and domain			
	Listening	Speaking	Reading	Writing
5	.46	.26	.50	.43
8	.35	.22	.40	.27
11	.28	.13	.35	.25

Table III-4 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, 31% of proficient ELs in grade 3 scored at level 3 or level 4 (proficient) in KAP ELA compared to 13% of proficient ELs in grade 7, who scored at level 3 or level 4. Proficient ELs at grade 10 had the lowest performance in KAP ELA and mathematics: only 4% of proficient ELs scored at level 3 or level 4 on KAP ELA and mathematics.

Table III-4. 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	469	31	52	17	471	56	34	9	—	—	—	—
4	666	32	59	9	668	27	57	16	—	—	—	—
5	504	17	52	31	507	16	49	35	508	28	47	25
6	192	18	43	39	193	14	42	44	—	—	—	—
7	223	13	40	47	225	16	53	32	—	—	—	—
8	220	8	45	46	222	10	31	59	222	9	25	66
10	248	4	40	56	248	4	24	72	—	—	—	—
11	—	—	—	—	—	—	—	—	104	12	24	64

### 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* (AAI, 2021a). A teacher survey collected an additional piece of evidence based on consequences of testing during the 2022 KELPA administration. [Appendix B](#) and [Appendix C](#) presented results of the teacher survey for select response and open-ended response questions. Responses to some survey questions indicated that most participating educators believed that the content of KELPA measured important English language proficiency knowledge, skills, and abilities (77%,  $n = 115$ <sup>2</sup>) and that expectations of ELs were aligned to knowledge and skills needed in the classroom (68%,  $n = 102$ <sup>3</sup>).

#### III.3.1 Teacher Survey

In the current document, the results in Table B-1 show that more than half (62%) of the participating educators who responded to the survey were teachers (i.e., classroom, Title 1, special education, English learners [ELs]) who administered KELPA. Many of these educators had 10 or more years of experience in ELA (62%), mathematics (55%), science (44%), and/or with ELs (46%; see Table B-3). They were well distributed

<sup>2</sup> 6% of participants did not respond to this survey question.

<sup>3</sup> 7% of participants did not respond to this survey question.



across different grades or grade bands (see Table B-2). Most educators (76%) believed that, overall, the English language support service provided to their students had been similar to a typical year before the COVID-19 pandemic (see Table B-4).

The percentage of educators who thought it somewhat easy or very easy to use Kite® Educator Portal ranged from 66% (e.g., upload student batch scores) to 90% (e.g., upload student responses for kindergarten or grade-1 writing). Uploading student batch scores and assigning raters are specific District Test Coordinator tasks and therefore had fewer valid responses ( $n = 49$  and  $n = 57$ , respectively). Refer to Table B-5 for educators' responses about user experience of other aspects of Educator Portal. Nearly all educators (94%–96%) agreed or strongly agreed that both the technology practice test and the KELPA practice tests familiarized students and teachers with the technologies, format, and procedures of the real tests (see Table B-6 and Table B-7).

Most educators agreed or strongly agreed that the training materials for scoring were helpful (88%) and that it was easy to access the audio files for speaking responses in the rater-training materials (83%; see Table B-8). For the survey question regarding the time the educators' districts spent on rater training before KELPA administration, 43% indicated that their districts spent one hour on rater training and 49% said their districts spent one and a half hours or more (i.e., up to six hours) on rater training. Refer to Table B-9 for a detailed distribution of the training hours.

Table B-10 shows that the vast majority of educators responded positively to questions about the utility of the [2021–2022 KELPA Examiner's Manual](#) (94%) and the *KELPA Test Administration and Scoring Directions* for both speaking and writing files (98%).

As shown in Table B-11, most educators agreed or strongly agreed that KELPA content measured important English language proficiency knowledge, skills, and abilities and expectations of ELs aligned to what is needed in the classroom (82%); that ELs classified as proficient based on KELPA are able to fully access grade-level academic content (83%) and those identified as not proficient cannot (75%); and that KELPA results provide useful information to future English for speakers of other languages (ESOL) services (78%).

The teacher survey included some open-ended questions to collect educators' feedback on a variety of topics. These topics included differences between English language support in the 2021–2022 academic year compared to a typical, prepandemic year; suggestions for additions and changes to Educator Portal; materials used by the district for rater training; and usage of KELPA test results.

Teachers described several differences between English language support in 2021–2022 compared to typical years (i.e., prepandemic), including high stress levels, staffing and support-personnel shortages, academic achievement gaps, higher levels of student absences, and negative instructional impacts because of masking requirements and remote learning. Teachers suggested adding current and previous scores and results to Educator Portal, including multilingual reporting options, as well as options to sort student tickets (i.e., student request notes used for on-campus communication). Teachers also suggested adding a search box, updating the navigation features, and including information on academic expectations and skills. Additionally, teachers wanted more details in the scoring rubrics, as well as a more streamlined scoring process. Teachers reported that various materials were used by their districts for rater training, including rubrics, sample responses, KELPA manuals and guides, and training videos. In response to a question about the utility of KELPA results (other than informing future ESOL services), teachers suggested various uses: to drive instruction, to determine student placement, and to determine the need for special education services.

Refer to Appendix C Responses to Open-Ended Teacher Survey Questions for more information about educators' responses.

## IV. Technical Quality—Other

This chapter provides updated evidence related to the technical quality of KELPA administered in 2022, 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* (AAI, 2021a).

### 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 the 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 2022 data. IRT reliability does not change by test sample, only by test form. Because the same test forms were used in 2020, 2021, and 2022, 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* (AAI, 2021a). For CTT reliability coefficients, the student-group reliabilities were also calculated. Indices of classification consistency and accuracy of different domain performance levels and 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* (AAI, 2021a).

#### 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 .79 to .97 across most grades or bands and domains. The exceptions were in kindergarten for reading (.68) and writing (.73). 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* (AAI, 2021a) 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	Coefficient $\alpha$			
	Listening	Speaking	Reading	Writing
K	.85	.93	.68	.73
1	.86	.93	.89	.82
2–3	.89	.93	.90	.86
4–5	.88	.94	.84	.84
6–8	.87	.95	.84	.86
9–12	.90	.97	.85	.81

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

Reliability estimates were also calculated by 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,  $\alpha = .77$ ) 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	.86	.85	.87	.86	.85
1	.85	.86	.85	.87	.85	.87	.86	.85
2–3	.88	.89	.88	.90	.88	.91	.89	.88
4–5	.87	.89	.87	.90	.87	.90	.88	.87
6–8	.88	.87	.87	.87	.87	.87	.86	.87
9–12	.89	.91	.90	.90	.90	.90	.89	.91
Speaking								
K	.93	.93	.93	.93	.93	.93	.93	.93
1	.93	.93	.93	.94	.93	.94	.93	.93
2–3	.92	.93	.92	.94	.92	.94	.92	.93
4–5	.94	.93	.93	.95	.93	.95	.93	.94
6–8	.96	.95	.95	.95	.96	.95	.96	.95
9–12	.97	.97	.97	.97	.97	.97	.97	.98
Reading								
K	.68	.68	.63	.77	.63	.76	.64	.69
1	.89	.89	.88	.90	.88	.91	.88	.89
2–3	.90	.90	.90	.91	.90	.91	.90	.90
4–5	.83	.84	.83	.85	.83	.85	.84	.82
6–8	.84	.84	.84	.85	.84	.85	.83	.83
9–12	.84	.86	.85	.85	.85	.85	.83	.86
Writing								
K	.73	.73	.71	.77	.72	.76	.71	.74
1	.81	.82	.81	.83	.81	.83	.82	.81
2–3	.86	.87	.86	.87	.86	.88	.86	.86
4–5	.83	.84	.82	.86	.83	.85	.84	.82
6–8	.87	.86	.86	.87	.86	.87	.85	.86
9–12	.81	.80	.80	.82	.80	.82	.77	.82

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

#### IV.1.2 Classification Consistency and Accuracy

When an assessment uses achievement or proficiency levels as the primary method 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 both 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.

The 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* (AAI, 2021a). 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 .68 to .98 across most cuts and grades or grand bands. Classification-accuracy indices for the KELPA domain tests ranged from .76 to .98 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	.92	.95	.90	.93	.77	.84
1	.94	.96	.89	.92	.80	.85
2	.97	.98	.92	.94	.86	.90
3	.98	.98	.95	.96	.89	.92
4	.96	.97	.95	.97	.85	.90
5	.97	.98	.95	.97	.85	.90
6	.95	.97	.94	.96	.83	.88
7	.95	.97	.93	.95	.84	.89
8	.95	.96	.93	.95	.84	.89
9	.92	.94	.90	.93	.87	.91
10	.94	.96	.92	.95	.89	.92
11	.93	.95	.92	.94	.85	.90
12	.94	.96	.93	.95	.86	.90
<b>Speaking</b>						
K	.92	.94	.89	.92	.83	.87
1	.97	.98	.93	.95	.76	.83
2	.97	.98	.94	.96	.81	.87
3	.97	.98	.96	.97	.81	.87
4	.98	.98	.96	.97	.85	.90
5	.97	.98	.96	.97	.77	.85
6	.97	.98	.95	.97	.84	.89
7	.97	.98	.95	.97	.82	.88
8	.97	.98	.96	.97	.77	.84
9	.96	.97	.96	.97	.92	.95
10	.97	.98	.96	.97	.92	.95
11	.97	.98	.96	.97	.89	.92
12	.97	.98	.96	.98	.88	.92

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>Reading</b>						
K	.68	.76	.84	.89	.93	.95
1	.86	.90	.87	.91	.92	.94
2	.86	.90	.89	.92	.90	.93
3	.91	.94	.91	.93	.87	.91
4	.91	.94	.85	.90	.82	.87
5	.90	.93	.84	.89	.81	.86
6	.94	.96	.84	.89	.84	.89
7	.91	.93	.85	.89	.82	.87
8	.91	.94	.85	.90	.77	.82
9	.85	.90	.87	.91	.90	.93
10	.86	.90	.85	.89	.86	.90
11	.86	.90	.85	.90	.86	.90
12	.87	.91	.84	.89	.83	.88
<b>Writing</b>						
K	.82	.87	.77	.83	.89	.93
1	.93	.95	.84	.89	.75	.80
2	.92	.94	.87	.91	.80	.86
3	.94	.96	.88	.92	.75	.79
4	.93	.95	.90	.93	.77	.83
5	.95	.97	.89	.92	.71	.78
6	.95	.97	.90	.93	.76	.82
7	.96	.97	.86	.90	.76	.81
8	.96	.97	.86	.90	.73	.79
9	.86	.90	.80	.86	.79	.85
10	.90	.93	.82	.87	.80	.85
11	.85	.89	.81	.86	.79	.85
12	.86	.90	.82	.87	.78	.84

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

#### IV.1.3 Interrater-Agreement Study

The purpose of the rater-agreement study is to provide 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* (AAI, 2021a) 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 rater-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 2022 KELPA scoring window (January 1–March 31, 2022). 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. Scores of record (used in score reporting) for operational scoring remained the same for all students regardless of whether a student was selected for second rating (i.e., the first score entered; refer to [Section IV.3.1.2 Educator Scoring](#) of the *2020 KELPA Technical Manual* [AAI, 2021a] for more information about how scores were entered.). 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	
	<i>Option 1</i>	<i>Option 2</i>
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 for the first rating: Users may select individual (i.e., scoring items individually) or paired/group (i.e., scoring items in pairs or a small group) scoring.
- Speaking scoring options for the first rating: Users may select simultaneous (i.e., scoring items in the moment that students are responding) or deferred (i.e., scoring items later by listening to the recordings) scoring.
- Designated scorer for the first rating: Default to user logged in; users may change name of scorer if scored by another user.
- Scoring method for the second rating: Users may select individual or paired/group scoring.
- Speaking scoring options for the second rating: Users may select simultaneous or deferred scoring.
- Designated scorer for the second rating: Default to user logged in; users may change name of scorer if scored by another user.



### IV.1.3.2 Sampling

A sample of students taking KELPA for the 2022 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 11% of registered students in grades 2–12 also was 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	34	132	485
1	40	134	474
2–3	46	188	744
4–5	42	172	596
6–8	47	103	709
9–12	50	79	815

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 and 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	No. of student responses per item	
	Writing: Individual scoring	Speaking: Combination of individual and deferred scorings
K	432–434	332–343
1	441–443	337–343
2–3	684–689	577–584
4–5	544–550	466–469
6–8	666–668	612–621
9–12	793–796	659–675

### IV.1.3.3 Raters

KELPA constructed responses are 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 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 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* (AAI, 2021a) 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 one another 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 also can measure agreement when three or more performance levels are used. *Weighted kappa* distinguishes between the numbers 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 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):

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

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

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 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 59% (grade band 6–8) to 83% (grade 1). The average percentage of exact-plus-adjacent agreement across items within 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	76	76	97	98
1	83	82	98	98
2–3	75	75	98	98
4–5	71	70	98	98
6–8	60	59	97	96
9–12	63	62	97	97

Table IV-8 summarizes agreement for speaking items. For speaking responses, the average percentage of exact agreement across items within 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 method—ranged from 63% (kindergarten) to 70% (grade band 9–12). The average percentage of exact-plus-adjacent agreement across items within grade or grade band—for overall, the individual scoring method, and the combination of individual and deferred scoring methods—was 94% 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	61	63	95	95	96
1	67	67	68	96	96	97
2–3	71	71	71	98	98	98
4–5	70	70	70	97	97	97
6–8	66	66	66	97	97	97
9–12	69	69	69	94	94	94

*Note.* Individual + deferred = combination of individual and deferred scoring methods.

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 grades 1–3 writing items. For both speaking and writing, lower grades (i.e., kindergarten through grade 3) had better agreement than higher grades. The only exception was that all speaking items in grades 9–12 had excellent agreement.

*Table IV-9. Summary of Quadratic Kappa Classifications*

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

#### IV.1.3.4.3 Summary

Individual scoring was the dominant scoring method for both writing and speaking items in 2022. 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 59% to 83% for writing responses and from 63% to 70% 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 94% or greater for speaking responses. Statistics for the quadratic-weighted kappa show that, for writing responses, raters had excellent agreement on 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 from kindergarten through grade band 6–8. The exception was that raters had excellent agreement on all speaking items in grade band 9–12. 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 2022 administration. For information about the procedures for scoring individual items, scoring the test as a whole, scaling, and specific quality-control process followed by the Achievement and Assessment Institute (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* (AAI, 2021a).

### IV.2.1 Operational Test Results

The number of students who took KELPA in 2022, 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 2022 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 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 2020 to 2022. The participation rate or tested rate for 2022 KELPA, computed by number of students tested divided by number of students enrolled, ranged from 83% to 99%, with the lowest participation rates in high school grades.

The participation rates for the 10 State Board of Education (SBOE) districts in 2022 are presented in Table IV-10 by grade or grade band. Kansas has 286 school districts that are separated into 10 SBOE districts. The participation rates (i.e., tested rates) ranged from 90% (SBOE districts 5, 7, 8, and 10 in grade band 9–12) to 100% (SBOE district 6 in grade band 4–5). 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 are the Kansas City, Kansas Public Schools district (part of SBOE district 1, whose average tested rate was 98% across grades and grade bands) and the Wichita Public Schools district (part of SBOE district 7, whose average tested rate was 97% across grades and grade bands). Both school districts are in SBOE districts that had remarkably high participation rates in elementary and middle schools but decreased participation rates in high schools. The decreased participation rates in high schools in these two SBOE districts are consistent with the dramatic enrollment drop from 2021 to 2022 in grades 10–12 reported in Table IV-20, indicating that the two largest school districts experienced a significant impact from the pandemic on both enrollment and participation rates.

Table IV-10. 2022 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	1,334	100 <sup>a</sup>	1,301	99	2,399	99	1,978	99	2,265	99	2,561	91
2	764	99	736	99	1,275	100	977	99	1,020	99	1,127	95
3	708	99	680	99	1,165	100	899	99	944	99	1,021	95
4	213	99	235	97	421	97	355	97	410	95	535	90
5	1,083	99	1,004	100 <sup>a</sup>	1,898	99	1,612	99	1,939	99	2,252	96
6	213	99	196	99	301	100 <sup>a</sup>	214	100	245	99	227	97
7	1,045	99	1,027	99	2,001	99	1,544	99	1,949	97	2,417	90
8	852	99	848	98	1,630	99	1,303	99	1,631	97	2,081	90
9	163	99	164	100	321	100	259	99	256	98	304	97
10	976	99	994	98	1,877	99	1,483	99	1,839	97	2,306	90

<sup>a</sup> Calculated as 100% because of rounding.

For all tested ELs, Table IV-11 shows the percentage of students in each demographic group by grade<sup>4</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>4</sup> Economically disadvantaged (ED) status is not shared with ATLAS 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 = 4,597)	1 (n = 4,436)	2 (n = 4,342)	3 (n = 3,884)	4 (n = 3,583)	5 (n = 3,061)	6 (n = 2,639)	7 (n = 2,619)	8 (n = 2,387)	9 (n = 2,736)	10 (n = 2,093)	11 (n = 1,878)	12 (n = 1,510)
<b>Race</b>													
Black	4.1	5.0	4.7	4.2	4.2	4.4	4.9	4.8	4.4	6.0	4.9	5.7	6.5
American Indian	6.4	6.1	7.1	8.2	7.2	7.9	8.9	9.8	11.0	11.4	12.9	15.7	19.1
Asian	12.1	1.3	1.3	9.9	8.9	8.3	7.6	7.0	7.2	7.0	7.2	7.0	8.1
NHPI	1.2	1.3	1.5	1.6	1.3	1.4	1.5	1.7	1.0	1.3	.7	.7	.8
White	76.2	77.3	76.3	76.0	78.4	78.0	77.1	76.7	76.4	74.3	74.3	7.9	65.6
<b>Hispanic</b>													
Yes	73.2	76.1	78.3	81.0	82.3	83.1	83.8	83.8	85.0	83.1	85.4	84.8	82.4
No	26.8	23.9	21.7	19.0	17.7	16.9	16.2	16.2	15.0	16.9	14.6	15.2	17.6
<b>SWD</b>													
Yes	28.3	27.8	28.4	33.3	34.1	35.0	36.6	35.3	38.4	34.4	34.4	28.0	3.4
No	71.7	72.2	71.6	66.7	65.9	65.0	63.4	64.7	61.6	65.6	65.6	72.0	69.6
<b>Gender</b>													
Female	48.8	47.8	46.8	47.1	46.7	43.7	43.9	43.6	42.4	43.5	42.4	42.7	44.1
Male	51.2	52.2	53.2	52.9	53.3	56.3	56.1	56.4	57.6	56.5	57.6	57.3	55.9

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<sub>10</sub>, P<sub>25</sub>, P<sub>50</sub>, P<sub>75</sub>, and P<sub>90</sub>, respectively. The differences between (a) P<sub>50</sub> and P<sub>25</sub> and (b) P<sub>75</sub> and P<sub>50</sub>, 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 symmetric in grade 11; negatively skewed in grades 2–4, 7; and 9–10, and positively skewed in other grades. For the speaking test, the distribution of scale scores was positively skewed in kindergarten and grade 4; distributions for other grades were skewed negatively. For the reading test, the distribution of scale scores was approximately symmetric in grades 9–10, positively skewed in grades 1–4 and 6–7, and negatively skewed in other grades. For the writing test, the distribution of scale scores was positively skewed in grades 1, 4, 6, and 10 and negatively skewed in other grades.

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

Grade	<i>M</i>	<i>SD</i>	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	518.86	165.45	0	354	421	492	589	695	1,000
1	498.87	14.59	0	343	421	480	573	626	1,000
2	483.99	156.97	0	328	391	475	541	605	1,000
3	574.07	209.76	0	365	453	541	605	1,000	1,000
4	513.69	178.81	0	349	411	491	535	611	1,000
5	549.85	193.40	0	362	432	491	611	1,000	1,000
6	477.24	116.48	0	347	414	478	552	615	1,000
7	506.44	137.11	152	347	432	510	552	615	1,000
8	539.18	163.33	0	347	432	510	615	725	1,000
9	461.93	154.14	0	303	360	455	506	622	1,000
10	503.30	17.02	0	327	394	477	547	622	1,000
11	515.22	179.56	0	338	407	477	547	622	1,000
12	537.35	188.81	0	338	421	506	622	1,000	1,000

*Note.* 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	<i>M</i>	<i>SD</i>	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	483.33	156.36	0	329	434	503	580	634	1,000
1	525.29	182.76	0	366	448	526	576	640	1,000
2	515.22	175.07	0	364	434	500	550	616	1,000
3	556.73	195.06	0	386	472	531	575	1,000	1,000
4	537.29	20.27	0	366	447	502	577	1,000	1,000
5	559.95	22.62	0	366	447	520	577	1,000	1,000
6	50.01	191.32	0	355	430	485	536	583	1,000
7	515.86	206.87	0	346	432.5	496	555	1,000	1,000
8	532.80	235.11	0	327	440	508	569	1,000	1,000
9	512.98	282.59	0	0	399	493	556	1,000	1,000
10	526.73	273.00	0	247.2	423	502	556	1,000	1,000
11	529.08	284.74	0	0	423	493	556	1,000	1,000
12	539.85	292.92	0	0	429	502	556	1,000	1,000

*Note.* 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	<i>M</i>	<i>SD</i>	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	48.98	125.97	0	363	399	463	522	618	1,000
1	482.37	128.49	0	369	393	451	548	648	1,000
2	46.32	121.54	0	347	377	441	516	606	1,000
3	537.06	16.87	0	377	428	516	606	673	1,000
4	484.34	131.55	0	329	388	465	557	665	1,000
5	522.36	144.82	0	358	422	521	602	665	1,000
6	471.41	112.88	0	336	390	463	541	628	1,000
7	498.77	128.23	0	355	407	485	579	628	1,000
8	523.04	133.05	0	355	424	511	579	699	1,000
9	448.96	106.93	0	338	377	439	502	594	1,000
10	479.03	114.08	0	359	393	469	542	631	1,000
11	486.19	116.19	0	359	393	485	566	631	1,000
12	505.48	12.44	0	359	424	502	566	631	1,000

Note. 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	<i>M</i>	<i>SD</i>	Min	P <sub>10</sub>	P <sub>25</sub>	P <sub>50</sub>	P <sub>75</sub>	P <sub>90</sub>	Max
K	497.04	146.46	0	342	401	499	554	637	1,000
1	498.92	164.13	0	336	400	464	588	691	1,000
2	454.84	124.72	0	298	368	465	523	580	1,000
3	515.29	129.43	0	355	434	523	580	687	1,000
4	48.07	126.46	0	319	418	479	563	600	1,000
5	517.53	133.10	0	367	437	532	600	649	1,000
6	487.96	135.30	0	340	410	471	557	596	1,000
7	507.25	147.92	0	340	428	496	557	652	1,000
8	542.30	172.58	0	353	448	525	596	652	1,000
9	434.69	127.10	0	273	357	448	511	555	1,000
10	473.94	121.96	0	321	411	468	532	585	1,000
11	486.65	125.31	0	340	411	490	555	632	1,000
12	507.03	127.83	0	357	429	511	555	632	1,000

Note. 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 56% (grade 4) across grades for speaking, from 9% (kindergarten) to 40% (grades 2–5) across grades for reading, and from 9% (kindergarten) to 41% (grade 10) across grades for writing.

Figure IV-1. 2022 Performance-Level Results for Listening

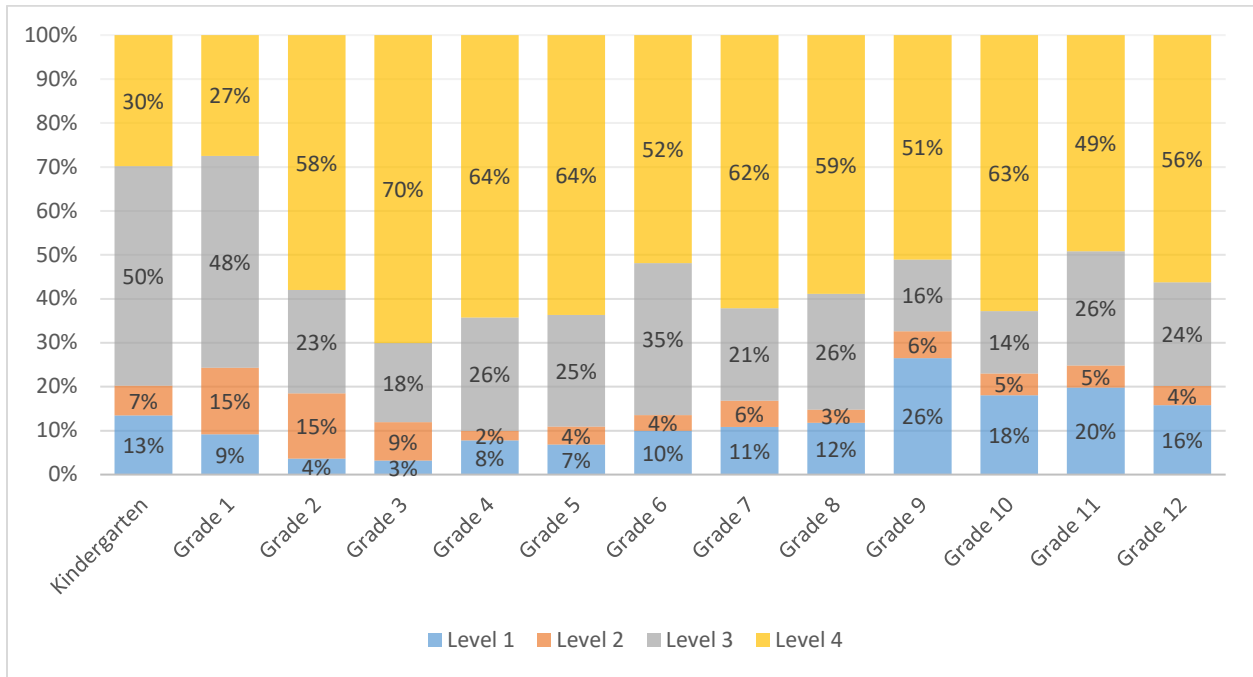


Figure IV-2. 2022 Performance-Level Results for Speaking

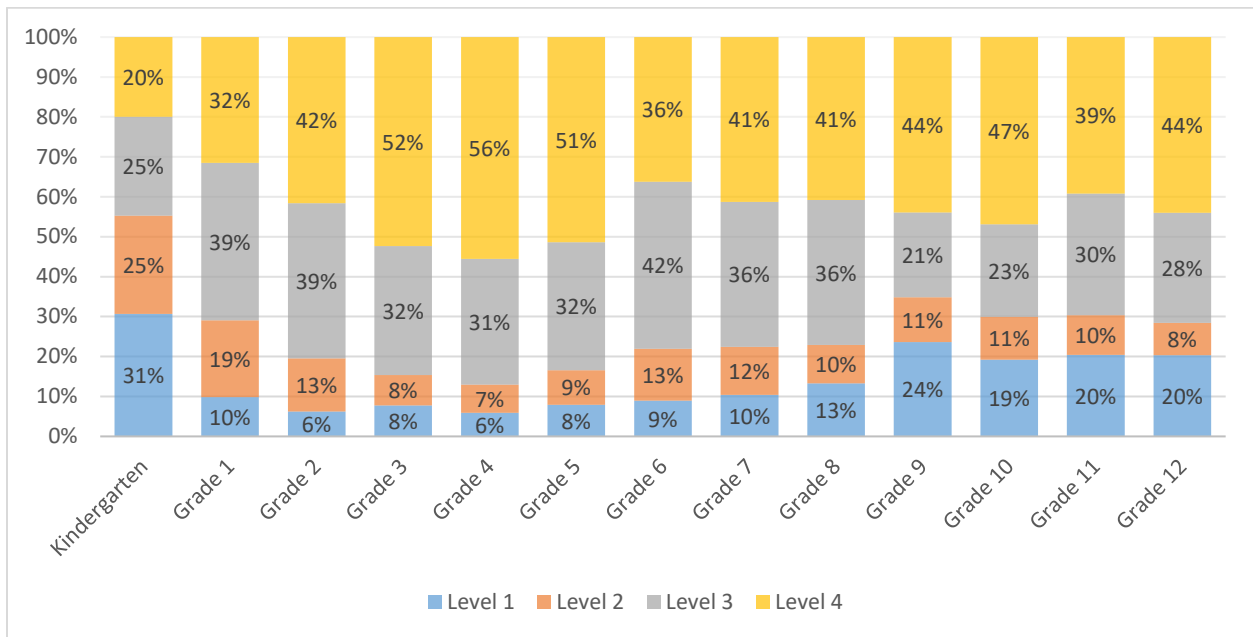


Figure IV-3. 2022 Performance-Level Results for Reading

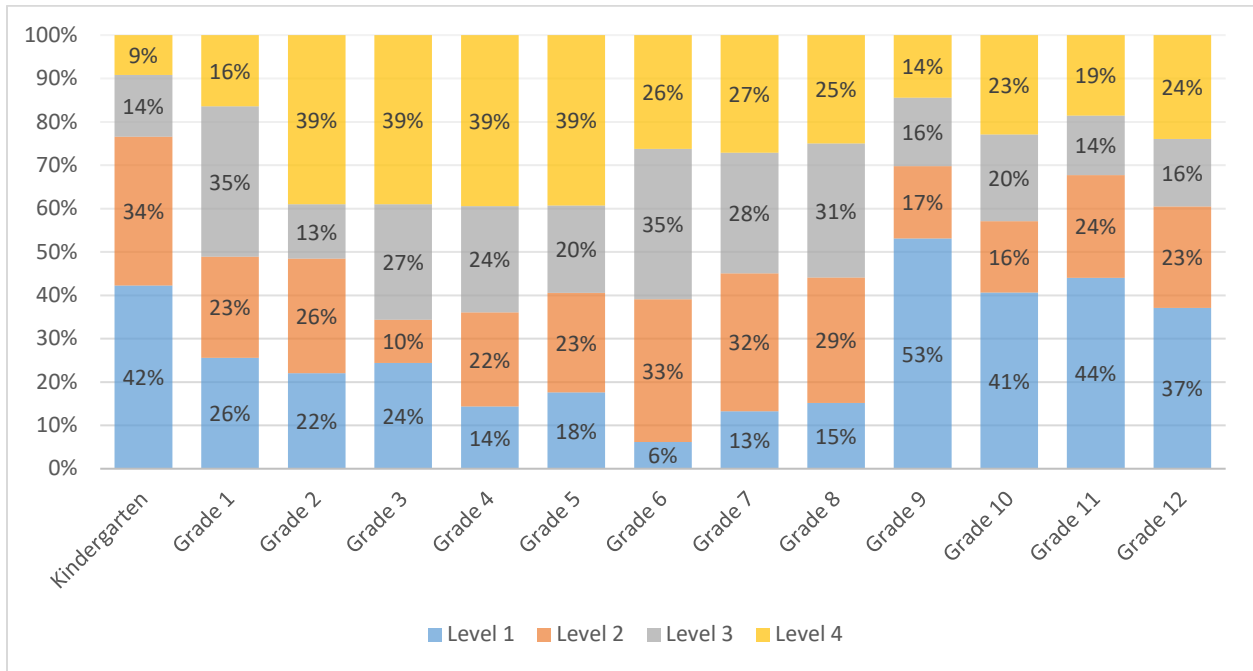
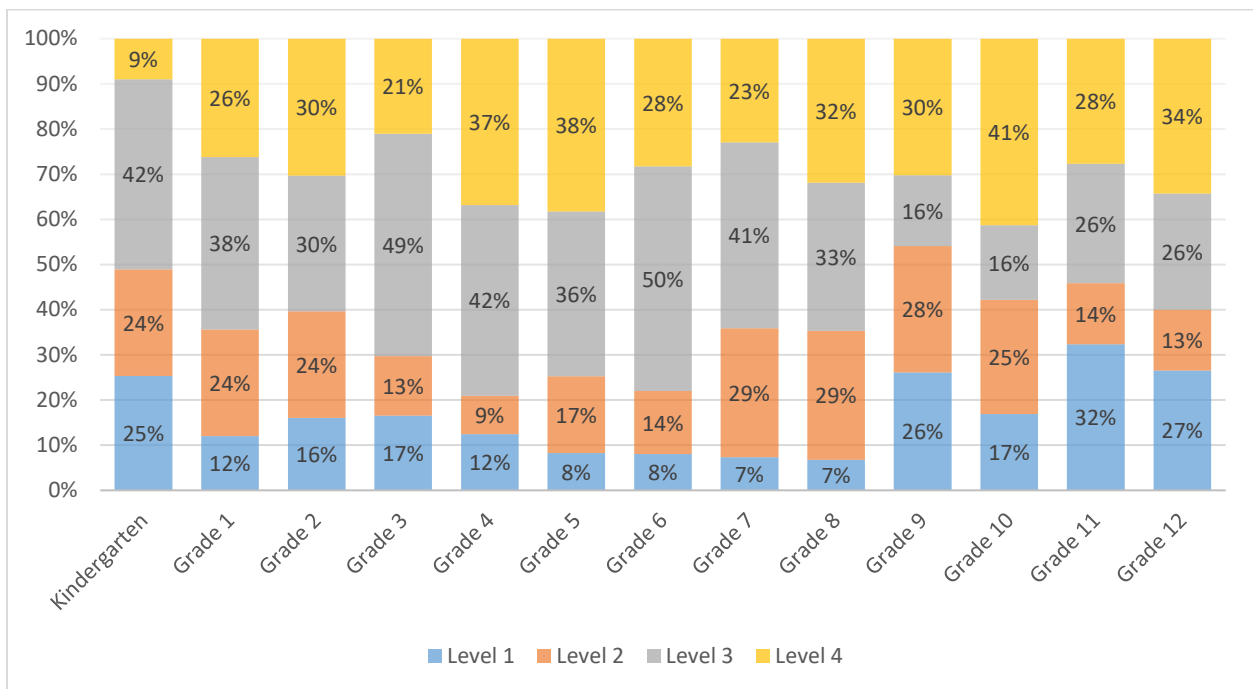


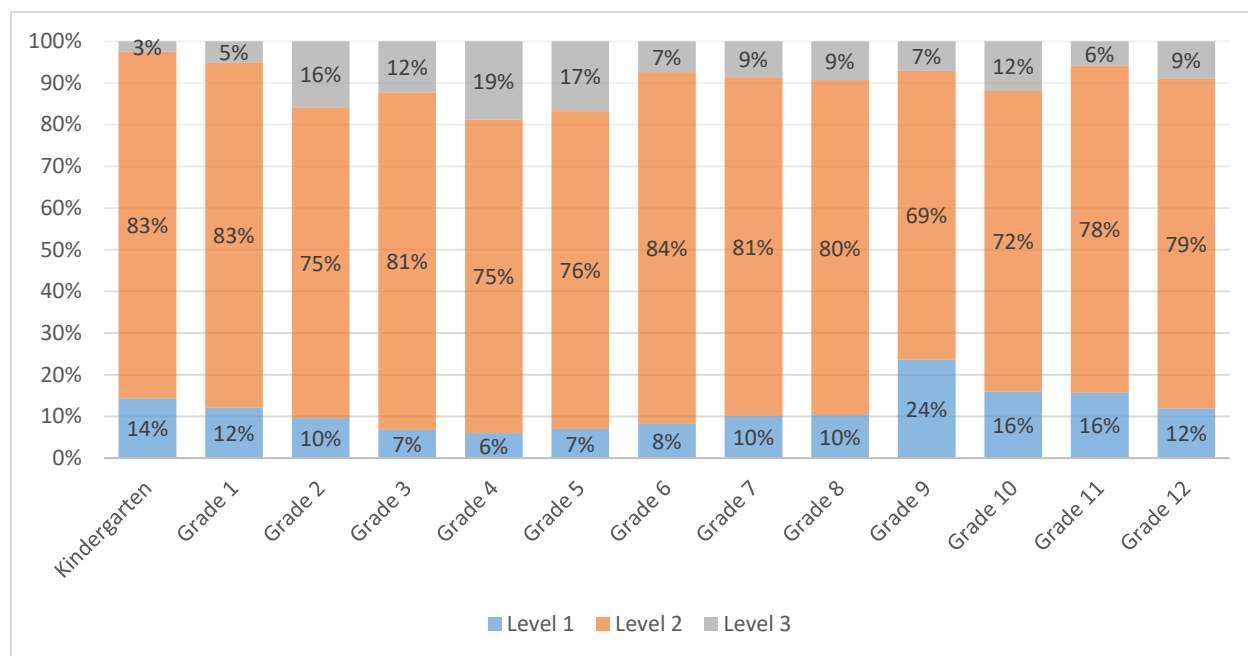
Figure IV-4. 2022 Performance-Level Results for Writing



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 proficiency level

2 (i.e., nearly proficient). The overall proficiency levels in 2022 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% (grade 6). Overall, the proficiency rates ranged from 3% (kindergarten) to 19% (grade 4). Kindergarten and grade 1 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 (2022 Administration)



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

Summaries of average scale scores by demographic groups<sup>5</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. However, NHPI students had the highest mean scores for the listening test in grades 1 and 6. For the speaking test, Black students had the highest mean scores in grade 7, White students had the highest mean scores in grade 3, and NHPI students had the highest mean scores in grades 1, 6, and 9–12. For the reading test, NHPI students had the highest mean score in grade 9, and White students had the highest mean scores in grade 12. For the writing test, NHPI students had the highest mean scores in grades 5, 6, and 10. 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 5 and 12 in listening; grade 12 in speaking, reading, and writing). Across all domains and grades, the mean scores of students without a disability were slightly higher than those of students with a disability, except for the speaking mean score in grade 9. For writing tests, the mean scores of female students were higher than those of male students in all grades. For the listening test, the mean scores of female students were higher than those of male students in most grades, except for grades 4–5 and 8. For the speaking test, the mean scores of female students were higher than

<sup>5</sup> Economically disadvantaged status is not shared with ATLAS to protect the privacy of students, so this student group is not included in the comparison.

those of male students in most grades, except for grade 8. For the reading test, the mean scores of the two groups were remarkably close or equal, with male students having slightly higher mean scores in grades 5, 8–9, and 11–12. These findings are similar to 2021 findings. Even when a test is carefully constructed with many considerations of fairness, differences may exist among student groups because of achievement gaps. Trend data comparing both the overall test results and results in each domain from 2020 to 2022 are provided in the next subsection.

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

Group	Grade																													
	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	M	SD		
Race																														
AI	495	149	482	140	480	142	576	204	507	180	530	181	477	134	502	137	525	152	460	167	494	176	503	179	530	194				
Asian	541	175	503	157	508	180	600	230	530	211	560	222	473	109	532	149	569	158	482	166	517	157	548	203	548	179				
Black	501	164	501	155	468	170	552	216	527	207	510	173	477	118	505	141	493	167	448	134	500	159	501	157	501	182				
NHPI	483	112	500	137	443	158	517	175	480	169	536	175	484	109	465	132	473	145	453	157	550	203	497	177	488	181				
White	519	165	499	138	483	155	572	209	512	173	552	192	477	115	504	134	541	163	461	152	503	170	516	177	543	189				
Hispanic																														
Yes	513	161	496	137	480	151	572	207	509	172	550	192	476	116	505	136	538	162	460	155	501	170	511	176	539	190				
No	535	177	507	152	497	176	584	222	535	207	549	202	481	120	514	143	545	172	472	152	520	169	542	199	532	185				
SWD																														
Yes	483	154	460	126	446	141	537	200	480	161	519	187	453	104	475	118	505	141	449	139	487	150	486	153	503	171				
No	533	168	514	143	499	161	592	212	531	185	567	195	491	121	524	144	561	172	469	161	512	179	527	188	552	194				
Gender																														
Female	532	167	515	141	494	158	587	216	507	168	537	177	484	119	513	137	538	161	464	150	509	165	516	169	549	189				
Male	507	163	484	139	475	156	562	203	520	188	560	205	472	114	501	137	540	165	460	158	499	174	515	187	528	188				

Note. 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	Grade																													
	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	M	SD		
Race																														
AI	454	173	497	159	508	164	531	185	513	186	529	210	493	196	506	213	524	230	521	280	537	255	515	269	536	297				
Asian	503	159	523	223	523	202	550	225	554	242	586	264	512	197	524	210	559	215	541	295	509	254	550	257	527	311				
Black	487	161	538	186	503	202	555	216	498	177	503	193	509	201	537	215	537	233	497	272	545	272	524	286	539	310				
NHPI	482	181	550	216	478	166	529	151	486	184	528	214	555	207	473	178	470	281	561	297	571	233	568	200	553	165				
White	482	153	525	177	515	171	559	192	541	197	562	217	496	186	516	207	532	237	510	283	524	278	530	291	542	288				
Hispanic																														
Yes	479	154	522	177	515	171	555	189	537	197	559	217	495	189	512	207	528	237	508	283	525	274	526	288	542	293				
No	494	162	535	199	515	190	563	219	539	214	566	237	526	203	534	208	558	224	538	279	540	267	548	269	532	294				
SWD																														
Yes	448	157	489	169	476	152	522	173	508	186	531	206	479	194	480	198	478	217	537	281	514	274	507	277	493	272				
No	497	154	539	186	531	181	574	203	552	206	576	226	512	189	535	209	567	240	500	283	534	272	538	287	560	299				
Gender																														
Female	495	152	545	189	537	185	574	202	556	210	582	232	510	206	525	219	530	245	527	292	552	276	534	297	555	305				
Male	472	160	507	175	496	164	542	187	521	189	543	210	492	178	509	196	535	227	502	274	508	269	525	276	528	283				

Note. 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	Grade																													
	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	M	SD		
Race																														
AI	466	107	457	103	448	112	527	155	473	132	514	148	466	121	482	120	514	128	448	107	472	118	484	117	499	125				
Asian	545	166	544	160	512	154	570	168	514	156	536	168	495	114	532	142	563	132	459	110	498	120	500	112	503	114				
Black	496	155	505	152	455	147	520	167	476	137	486	132	472	122	510	161	464	120	436	100	454	106	465	106	456	113				
NHPI	463	112	479	99	430	90	519	131	444	108	517	125	474	147	484	134	490	104	461	129	487	103	477	124	478	72				
White	471	115	475	121	455	114	533	160	483	128	523	142	469	110	496	124	524	133	449	107	479	113	486	116	511	120				
Hispanic																														
Yes	470	115	471	118	452	112	533	159	480	128	522	142	468	112	495	125	522	132	448	107	478	113	486	117	509	120				
No	512	148	519	153	490	147	556	166	504	147	526	158	487	119	521	143	530	137	455	107	487	119	489	110	491	120				
SWD																														
Yes	458	120	454	115	427	109	498	148	444	122	489	141	430	96	458	115	485	120	436	95	462	107	465	107	480	110				
No	490	127	494	132	473	124	557	164	505	132	540	144	495	115	521	130	547	136	456	112	488	117	495	119	517	123				
Gender																														
Female	484	125	488	127	467	120	544	163	484	130	519	134	474	114	504	128	517	128	448	102	480	110	482	108	504	107				
Male	478	127	477	129	454	122	531	159	484	133	525	153	470	112	495	128	527	136	450	110	478	117	489	122	506	130				

Note. 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	Grade																											
	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	M	SD
Race																												
AI	481	137	463	136	444	113	499	132	464	129	506	139	489	140	511	150	539	167	436	117	477	122	479	134	500	149		
Asian	568	176	564	201	513	153	556	157	531	170	548	172	517	148	542	163	601	183	455	157	502	130	507	116	530	124		
Black	514	175	518	185	440	129	495	134	472	114	485	135	466	143	490	159	499	165	415	118	451	119	449	108	442	110		
NHPI	495	118	495	143	443	127	539	112	459	134	551	105	526	176	495	130	511	142	428	190	503	120	474	86	474	88		
White	486	137	492	158	449	119	510	124	476	119	517	127	485	131	505	146	540	172	435	125	472	121	490	126	513	122		
Hispanic																												
Yes	486	137	488	156	447	117	509	125	474	120	515	129	485	132	504	144	537	170	433	123	471	121	485	125	509	128		
No	527	166	533	185	484	145	544	145	510	148	531	152	503	149	527	165	571	187	443	145	490	129	493	127	500	128		
SWD																												
Yes	469	140	461	149	411	114	475	124	447	128	478	122	450	117	472	132	503	147	422	106	461	112	462	113	479	103		
No	508	148	514	167	472	125	535	128	497	122	539	134	510	140	527	153	567	183	442	136	481	126	496	128	519	136		
Gender																												
Female	507	146	512	169	465	122	524	127	490	124	535	137	509	148	527	156	558	185	455	126	497	127	511	131	531	133		
Male	488	147	487	159	446	126	507	131	471	128	504	128	472	122	492	139	531	162	419	125	457	116	468	117	488	121		

Note. AI = American Indian; NHPI = Native Hawaiian and Pacific Islander; SWD = students with disabilities.

## IV.2.2 Trend Data

The 2022 KELPA administration was the third administration of the new KELPA aligned with the [2018 Standards](#). The next subsections present changes in enrollment data and performance-level distributions from 2020 to 2022.

### IV.2.2.1 Comparison of Enrollment

Because of the impact of the COVID-19 pandemic on the 2020–2021 academic school year, the enrollment and test-participation rates decreased in each grade from 2020 to 2021 (see Table IV-20). For the 2022 administration, 40,826 students were enrolled, and 39,765 students tested; the overall participation rate was 97%. Participation rates across grades ranged from 83% (grade 12) to 99% (kindergarten through grade 4). Compared to the 2021 administration, the enrollments in 2022 increased for elementary grades (K–5) and decreased for grades 10–12.

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

Grade	2020			2021			2022			% enrollment change (2021 to 2022) <sup>a</sup>
	No. enrolled	No. tested	% participation	No. enrolled	No. tested	% participation	No. enrolled	No. tested	% participation	
K	4,614	4,522	98	4,305	4,090	95	4,638	4,597	99	8
1	4,619	4,573	99	4,434	4,212	95	4,471	4,436	99	1
2	4,734	4,734	100	4,336	4,119	95	4,376	4,342	99	1
3	4,051	4,051	100	3,926	3,730	95	3,929	3,884	99	0
4	3,829	3,791	99	3,536	3,359	95	3,623	3,583	99	2
5	3,242	3,210	99	3,041	2,889	95	3,114	3,061	98	2
6	2,809	2,809	100	2,724	2,452	90	2,692	2,639	98	-1
7	2,663	2,636	99	2,538	2,310	91	2,684	2,619	98	6
8	2,755	2,727	99	2,480	2,207	89	2,424	2,387	98	-2
9	3,110	3,079	99	2,551	2,092	82	2,844	2,736	96	11
10	3,129	3,066	98	2,495	1,996	80	2,205	2,093	95	-12
11	2,830	2,773	98	2,373	1,780	75	2,003	1,878	94	-16
12	2,179	2,092	96	2,094	1,361	65	1,823	1,510	83	-13
Total	44,564	44,063	99	40,834	36,597	90	40,826	39,765	97	0

<sup>a</sup> Positive values indicate the increased percentage; negative values indicate the decreased percentage.

#### 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 2020 through 2022 by domain and grade. From 2021 to 2022, for listening, the level-4 percentages stayed the same in grades 2 and 7–8, increased in grades 3 and 6, but decreased in most grades. For speaking from 2021 to 2022, the level-4 percentages stayed the same in kindergarten and grades 9–10, increased slightly in grades 1–4 and 7, and decreased slightly in other grades. For reading from 2021 to 2022, the level-4 percentages stayed the same in grade 5, increased in grades 3–4, and decreased in the other grades. For writing from 2021 to 2022, the level-4 percentages stayed the same in grades 4 and 7, slightly increased in grades 1 and 6, and slightly decreased in other grades.

Figure IV-6. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Listening

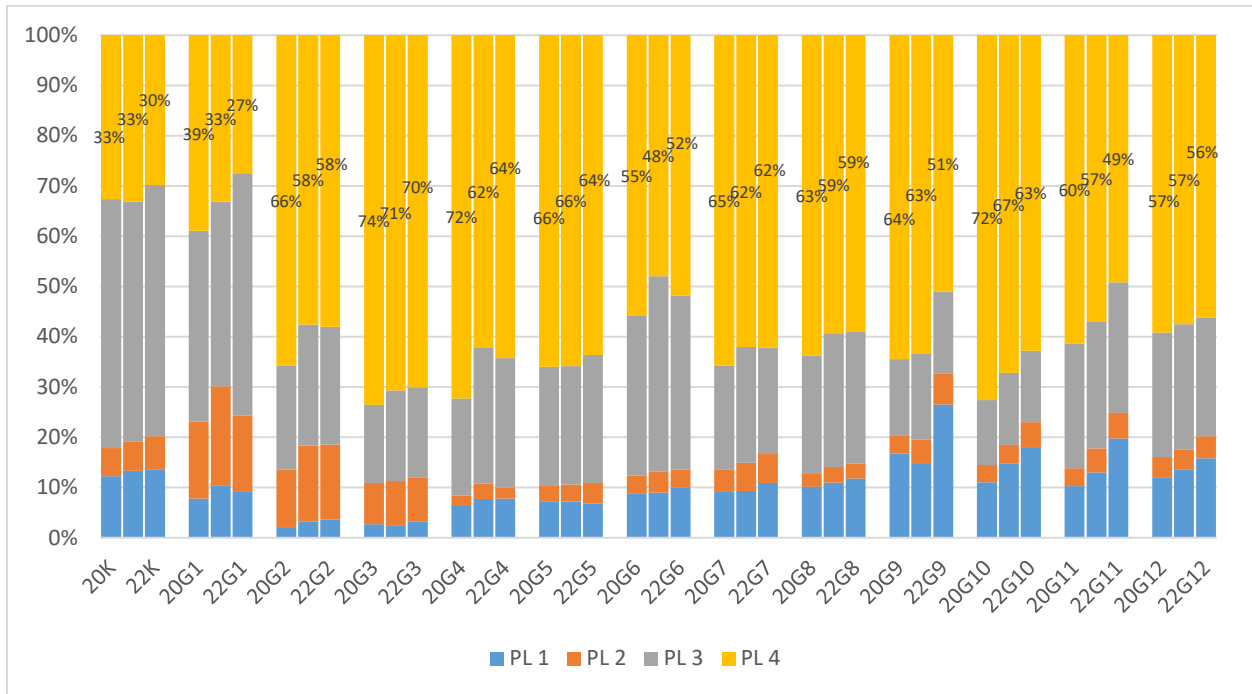


Figure IV-7. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Speaking

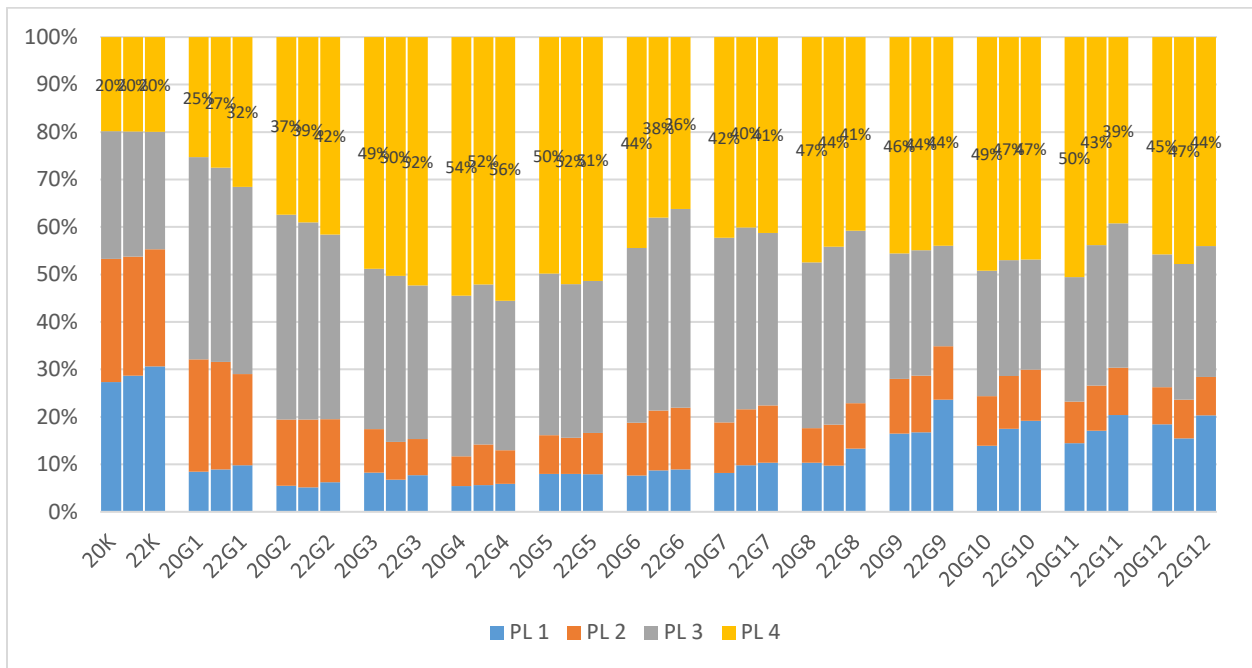


Figure IV-8. Comparison of Performance-Level (PL) Results From 2020 Through 2022 for Reading

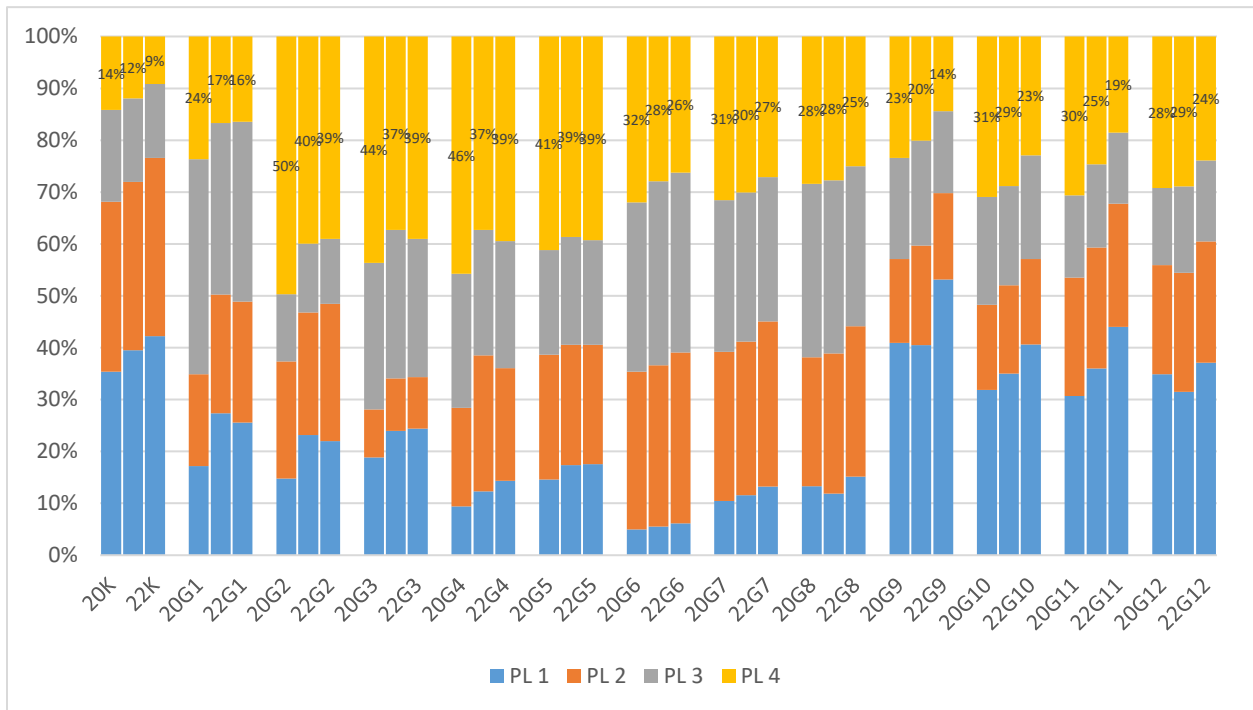
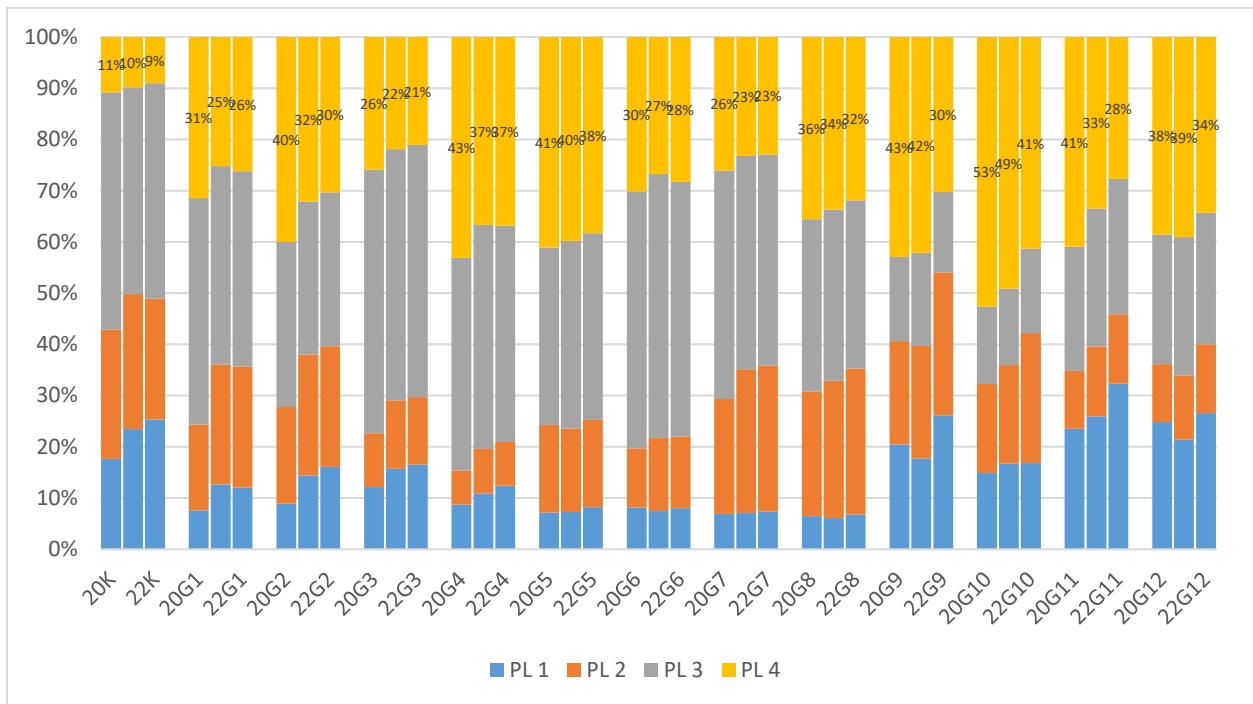
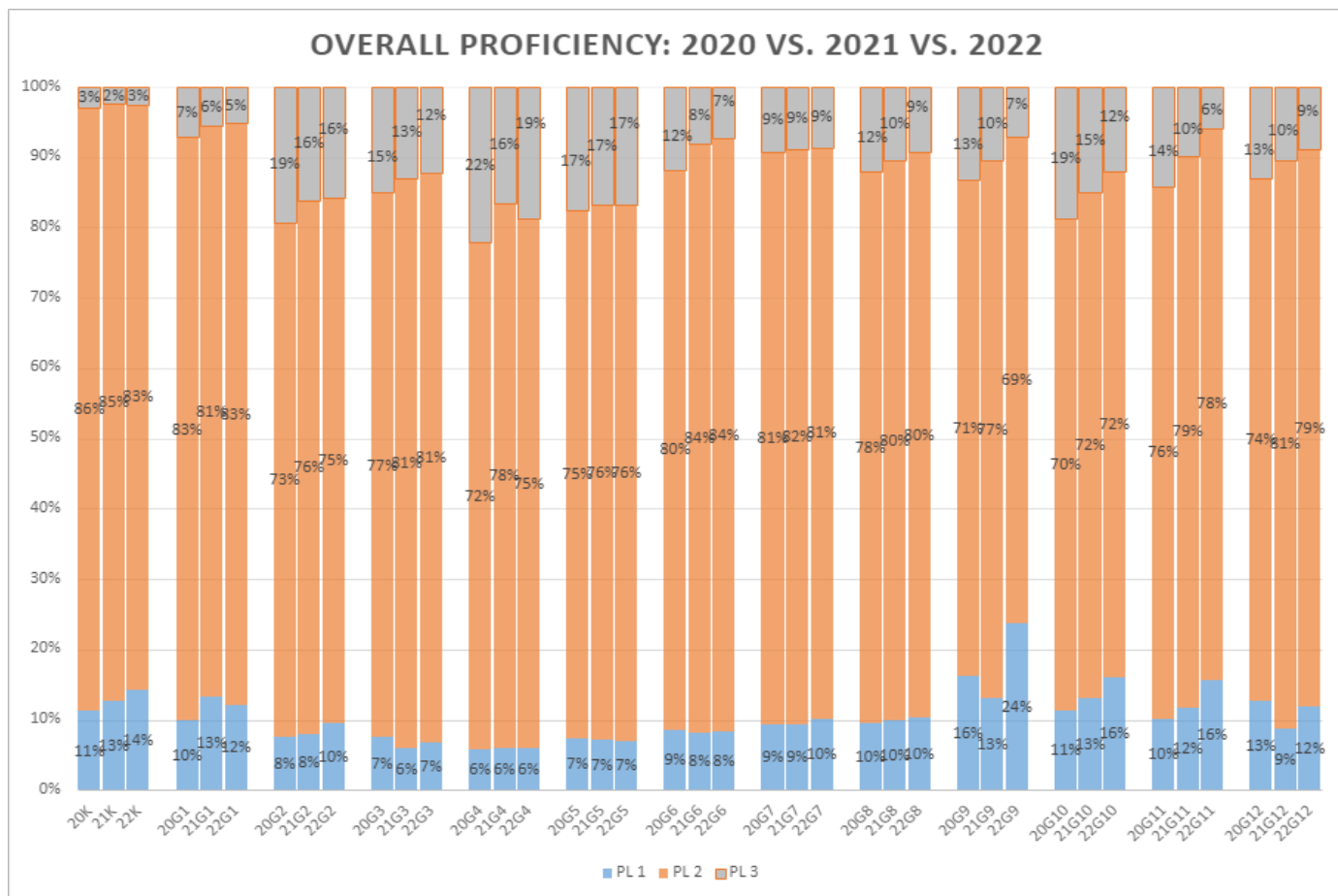


Figure IV-9. Comparison of Performance-Level (PL) Results from 2020 Through 2022 for Writing



The trend of the overall proficiency rates is provided in Figure IV-10. From 2021 to 2022, the overall proficiency rates stayed the same for grades 2, 5, and 7; increased slightly in kindergarten and grade 4 (1% and 3%, respectively); and decreased in other grades (from 1% to 4%). The proficiency rates in grades 5 and 7 stayed the same for all three years from 2020 to 2022.

Figure IV-10. Comparison of Overall Performance -Level (PL) Results From 2020 Through 2022



### IV.3 Ongoing Program Improvement

This section summarizes the ongoing improvements to KELPA.

#### IV.3.1 Enhanced Rater-Training Materials Development

The KELPA rater-training materials were redone for the 2021 administration to provide new prompts and exemplar student responses to one operational CR item per grade or grade band in speaking and writing. Since then, these materials have been expanded to cover all CR items for the 2023 administration. The purpose of the updated materials is to provide training materials that support educators in applying rubrics to specific prompts. For detailed information, refer to Section II.2.2 Development of Rater-Training Materials of the current manual.

### IV.3.2 Domain-Score Exemption

In certain situations, students may be exempt from taking a domain test. Unusual circumstances codes are available in Educator Portal, which allow school districts to manage test exemptions; these codes were enhanced to include KELPA domain exemptions for the 2021–2022 administration. Domain-exemption requests were reviewed 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 is determined by speaking, reading, and writing domain performance, and students are considered proficient overall if they score at level 4 in the speaking, reading, and writing domains. Table IV-21 shows the number of students exempted from testing by domain and grade for the 2022 administration. Speaking was the most likely domain in which students were exempted from testing (15 students across grades), and writing was the least likely domain (one student).

*Table IV-21. Number of Students Exempted From Testing by Domain and Grade*

Grade	Listening	Speaking	Reading	Writing
K	2	7	0	1
1	3	2	0	0
2	3	3	0	0
3	0	0	1	0
4	0	1	1	0
5	0	0	2	0
6	0	0	1	0
7	0	0	1	0
8	0	1	1	0
9	0	0	0	0
10	0	1	0	0
11	0	0	0	0
12	0	0	0	0
Total	8	15	7	1

### IV.3.3 Incident Response

A Kansas Assessment Program (KAP) system-wide incident response manual, which is applicable to the KELPA program, was finalized during the 2022 test administration. The purpose of the incident response manual is to guide investigative efforts of AAI staff for potential KAP or KELPA testing incidents. This response plan outlines the steps for managing and addressing any item-level or test-level incidents to remedy the effects and properly document relevant information.

## V. Inclusion of All Students

This chapter provides a summary of the frequency of accommodations used in 2022 KELPA administration and 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](#) of Chapter V 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>6</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 or 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* (AAI, 2021a) for guidelines applied 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 PNP submitted by teachers determine the

---

<sup>6</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.

availability of test accommodations for individual students. Table V-1 presents the number of students who took KELPA in Kansas in 2022 and had PNP accommodations. The summary in the table shows no accommodations were requested for kindergarten; two students in grade 1 and two in grade band 2–3 used whole screen magnification; 43 students in grade band 4–5, 25 students in grade band 6–8, and 102 students in grade band 9–12 used various accommodations. The most frequent accommodation (i.e., 116 students) was auditory calming, which provides relaxing, peaceful background music while a student takes the test. The second-most frequent accommodations (i.e., 24 students each) were color contrast and whole screen magnification.

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

Grade or grade band	No. of students using accommodation						
	Auditory calming	Color contrast	Color overlay	Masking	Reverse contrast	Switches	WSM
K	0	0	0	0	0	0	0
1	0	0	0	0	0	0	2
2–3	0	0	0	0	0	0	2
4–5	35	4	0	0	0	0	4
6–8	5	10	3	1	0	0	6
9–12	76	10	1	0	1	4	10
Total	116	24	4	1	1	4	14

*Note.* WSM = whole screen magnification.



## VI. Academic Achievement Standards and Reporting

The 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 of setting 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* (AAI, 2021a). Because there were no updates to anything related to standard setting or performance level during the 2021–2022 school year, this chapter briefly updates information about student score reports.

### VI.1 Reporting

The 2022 KELPA testing window ended on March 11, 2022, and the scoring window closed on March 31, 2022. KELPA student reports were made available to all school districts on April 21, 2022, and in the Parent Portal on April 28, 2022.

#### VI.1.1 Student Reports

Performance levels for listening, speaking, reading, and writing were used to determine overall proficiency level, which is defined by KSDE. To be considered proficient (i.e., level 3 on overall proficiency) and eligible to exit the 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 2021 and 2022 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 2022 KELPA student report kept the same format and information used in the 2021 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 on 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 EL proficiency.

## 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>
- 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. 2022 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.

The purposes of this survey are to provide feedback on teachers' and test administrators' testing experience with KELPA and to provide feedback on user experience with the technology. All responses are confidential, and results will be reported only for 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. You may perform many roles in your district, please select the roles that best describe you. [Please select all that apply.]
  - Building Test Coordinator (BTC)
  - Building User (BU)
  - District Superintendent
  - District Test Coordinator (DTC)
  - District User (DU)
  - Proctor
  - Program administrator
  - State Assessment Administrator
  - Teacher administering KELPA
  - Teacher not administering KELPA
  - Technology Director/Coordinator
  - Support staff
2. If you administered KELPA, for which grades/grade bands did you administer KELPA this year? [Please select all that apply.]
  - Kindergarten
  - 1
  - 2–3
  - 4–5
  - 6–8
  - 9–12

3. Please indicate your number of years of K–12 educational experience in each of the following areas.
- English language arts \_\_\_\_\_
- Mathematics \_\_\_\_\_
- Science \_\_\_\_\_
- English learners \_\_\_\_\_

**II. Learning and Instruction in 2021–2022**

1. Please rate the following statements about learning and instruction in 2021–2022.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
Overall, the English language support service provided to my students has been similar to a typical year before the COVID-19 pandemic.					

2. If you indicated that the English language support provided to your students this year is different from a typical year before the pandemic, please describe the main differences. [Open-ended response]

**III. Technology**

The following questions are about your use of Kite® Educator Portal and Student Portal.

1. Educator Portal is used to manage data for KELPA. Please rate how easy or hard it was to do the following in Educator Portal this year.

	Very Hard	Somewhat Hard	Somewhat Easy	Very Easy	Not Applicable
Navigate the site.					
Upload student responses for kindergarten or grade-1 writing.					
Upload batch student scores.					
Manage second scoring.					
Assign raters (as a District Test Coordinator, DTC).					

2. Please indicate your level of agreement or disagreement with the given statement on the Technology Practice Test.

Technology practice test...	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
Familiarized students and teachers with the procedures for answering different types of technology-enhanced items.					

3. Please indicate your level of agreement or disagreement with each given statement on the KELPA Subject-Oriented Practice Tests.

Items on KELPA Subject-Oriented practice tests...	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
Familiarized students and teachers with the assessment format.					
Familiarized students and teachers with the procedures for responding to different types of KELPA items.					

4. 5. What data would you like to see added to Educator Portal? [*Open-ended response*]
5. If you could change one thing about Educator Portal, what would you change? [*Open-ended response*]
6. Is there anything we should change in Educator Portal to make scoring easier? [*Open-ended response*]

#### IV. Scoring

1. The rater-training materials were helpful in applying rubrics for scoring students' responses to speaking items.

- strongly disagree
- disagree
- agree
- strongly agree
- not applicable

2. The rater-training materials were helpful in applying rubrics for scoring students' responses to writing items.

- strongly disagree
- disagree
- agree
- strongly agree
- not applicable

3. It is easy to access the audio files for speaking responses in the rater-training materials.

- strongly disagree
- disagree
- agree
- strongly agree
- not applicable

4. Please indicate how much time (in hours) your district spent on rater training before KELPA administration.

\_\_\_\_\_ hour(s)

5. Please list materials your district used for rater training. [*Open-ended response*]

**V. Resources**

1. Please rate the following statements about KELPA support materials.

	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Not Applicable
The <i>2021–2022 KELPA Examiner’s Manual</i> was useful and helpful.					
The <i>KAP Practice Test Guide for Educators 2021–2022</i> was useful and helpful.					
<i>Kite Educator Portal Manual for Test Coordinators</i> was useful and helpful.					
The KELPA Test Administration and Scoring Directions for speaking files were helpful.					
The KELPA Test Administration and Scoring Directions for writing files were helpful.					

*[If you are familiar with the performance of your English learner (EL) on KELPA, please respond to the following six questions.]*

## **VI. Utility of KELPA**

1. KELPA content measured important English language proficiency knowledge, skills, and abilities.
  - strongly disagree
  - disagree
  - agree
  - strongly agree
  
2. KELPA content measured expectations of ELs that were aligned to what is needed in the classroom.
  - strongly disagree
  - disagree
  - agree
  - strongly agree
  
3. In general, English learner students classified as Proficient based on KELPA are able to fully access grade-level academic content.
  - strongly disagree
  - disagree
  - agree
  - strongly agree
  
4. In general, English learner students classified as Not Proficient based on KELPA are not able to fully access grade-level academic content without the use of English for ESOL services.
  - strongly disagree
  - disagree
  - agree
  - strongly agree
  
5. KELPA test results provide useful information to future ESOL services.
  - strongly disagree
  - disagree
  - agree
  - strongly agree
  
6. If KELPA test results are or will be used for purposes other than informing future ESOL services, please indicate and/or describe. [*Open-ended response*]

## Appendix B. Summary Results of Teachers' Responses to Survey Questions <sup>7, 8</sup>

*Table B-1. Responses About Teacher's Role Relating to KELPA (N = 149)*

Role	N	%
Building Test Coordinator (BTC)	53	36
Building User (BU)	22	15
District Superintendent	0	0
District Test Coordinator (DTC)	8	5
District User (DU)	8	5
Proctor	22	15
Program Administrator	9	6
State Assessment Administrator	1	< 1
Teacher administering KELPA	93	62
Teacher not administering KELPA	2	1
Technology Director/Coordinator	1	< 1
Support staff	7	5

*Note.* The total count is greater than 149 because participants were asked to select all roles that might describe them.

*Table B-2. Distribution of Test Administrators by Grade or Grade Band (N = 149)*

Grade or grade band	% of respondents
K	20
1	19
2–3	20
4–5	18
6–8	13
9–12	10

*Table B-3. Educators' Professional Experience in Years (N = 149)*

Years	Experience with (%)			
	English language arts	Mathematics	Science	English learners
0–2	24	35	45	19
3–5	5	4	4	18
6–9	9	6	7	17
10 or more	62	55	44	46

<sup>7</sup> Percentages in the tables may not sum to 100% because of rounding.

<sup>8</sup> Blank and “not applicable” responses were excluded from sample count N.



*Table B-4. Educators' Responses About Learning and Instruction in 2021–2022 (N = 139)*

Statement	Disagree	Somewhat disagree	Somewhat agree	Agree
Overall, the English language support service provided to my students has been similar to a typical year before the COVID-19 pandemic.	6%	12%	22%	60%

*Table B-5. Educators' Responses About User Experience of Kite Educator Portal*

Task	N	Very hard	Somewhat hard	Somewhat easy	Very easy
Navigate the site	146	2%	12%	32%	54%
Upload student responses for kindergarten or grade-1 writing	82	2%	7%	28%	62%
Upload batch student scores	57	11%	23%	26%	40%
Manage second scoring	101	5%	10%	27%	58%
Assign raters (as a DTC)	49	4%	27%	38%	31%

Note. DTC = district test coordinator

*Table B-6. Educators' Responses About KELPA Technology Practice Test (N = 121)*

Technology practice test . . .	Strongly disagree	Disagree	Agree	Strongly agree
Familiarized students and teachers with the procedures for answering different types of technology-enhanced items.	3%	3%	58%	37%

*Table B-7. Educators' Responses About KELPA Subject-Oriented Practice Tests*

Items on KELPA Subject-Oriented practice tests . . .	N	Strongly disagree	Disagree	Agree	Strongly agree
Familiarized students and teachers with the assessment format.	112	2%	2%	57%	39%
Familiarized students and teachers with the procedures for responding to different types of KELPA items.	115	2%	4%	53%	41%

*Table B-8. Educators' Responses About KELPA Scoring*

Statement	N	Strongly disagree	Disagree	Agree	Strongly agree
The rater-training materials were helpful in applying rubrics for scoring students' responses to speaking items.	145	5%	8%	56%	32%
The rater-training materials were helpful in applying rubrics for scoring students' responses to writing items.	145	5%	8%	58%	30%
It is easy to access the audio files for speaking responses in the rater-training materials.	140	6%	11%	57%	26%

*Table B-9. Respondent's Report for Amount of Time District Spent on Rater Training Before KELPA Administration (N = 104)*

No. of hours	% of respondents
0	7
0.25	1
0.5	1
1	43
1.5	4
2	21
3	13
4	9
5	1
6	1

*Table B-10. Educators' Responses About KELPA Support Materials*

Statement	<i>n</i>	Disagree	Somewhat disagree	Somewhat agree	Agree
The 2021–2022 KELPA Examiner’s Manual was useful and helpful.	146	1%	5%	13%	81%
The KAP Practice Test Guide for Educators 2021–2022 was useful and helpful.	115	1%	4%	18%	76%
The Kite Educator Portal Manual for Test Coordinators was useful and helpful.	112	2%	4%	14%	81%
The KELPA Test Administration and Scoring Directions for speaking files were helpful.	143	1%	1%	20%	78%
The KELPA Test Administration and Scoring Directions for writing files were helpful.	142	1%	2%	18%	80%

*Table B-11. Educators' Responses About the Utility of KELPA*

Statement	<i>n</i>	Strongly disagree	Disagree	Agree	Strongly agree
KELPA content measured important English language proficiency knowledge, skills, and abilities.	127	4%	14%	73%	9%
KELPA content measured expectations of ELs that were aligned to what is needed in the classroom.	127	4%	22%	65%	9%
In general, English learner students classified as Proficient based on KELPA are able to fully access grade-level academic content.	116	6%	11%	66%	17%
In general, English learner students classified as Not Proficient based on KELPA are not able to fully access grade-level academic content without the use of English for ESOL services.	122	6%	19%	63%	12%
KELPA test results provide useful information to future ESOL services.	119	5%	17%	64%	14%

*Note.* EL = English learner; ESOL = English for speakers of other languages.

## Appendix C. Responses to Open-Ended Teacher Survey Questions

The 2022 KELPA teacher survey used an open-ended format to ask educators for feedback and opinions on topics, such as the usability of the Educator Portal and the usage of KELPA test results. Some of these questions <sup>9</sup> and a summary of responses follow.

### II. Learning and Instruction in 2021–2022

**Q2.** *If you indicated that the English language support provided to your students this year is different from a typical year before the pandemic, please describe the main differences.*

- Teachers described high levels of stress among students, families, and staff, with more than one teacher noting escalations in student behavior.
- Teachers described staffing and personnel issues, including frequent staff absences, staff resignations, not enough qualified English for speakers of other languages (ESOL) staff to work with students, and a profound shortage of substitute teachers and paraprofessionals who are ESOL-endorsed, which limited the support schools were able to provide to students and contributed to inconsistency in class routines.
- Teachers described how many students are behind academically and “playing catch up,” leading teachers to scaffold more while noting a larger academic gap between these students and their peers.
- Teachers described more-frequent student absences, and one teacher noted that the number of EL students infected with COVID was much higher than in the general student population.
- Teachers noted that mask requirements and virtual schooling negatively affected instruction in English pronunciation and phonetics, particularly for younger students and shy students with difficulty speaking in front of teachers and peers.

“Students have not received consistent support. We are putting a broken puzzle (educational opportunity) back together. The pieces (students) have changed, society has changed, and we (educators) never had an opportunity to reflect on what worked well and what the gaps in opportunity and inequities were. There have been basic needs (food, shelter, social/emotional) that have had to have priority over direct, effective, and strategic English language support. We have had less staff and inconsistent guidance and support from administration. We didn’t start from where we left off but where the students ‘should be’ according to grade level. As if nothing disrupted their education.”

### III. Technology

**Q5.** *What data would you like to see added to Educator Portal?*

- Teachers wanted to see scores and results.
- Teachers wanted to see multilingual reports and results for families.
- Teachers wanted to see a portal that is easier to navigate.
- Teachers wanted to see student tickets sorted by teacher and hour; one teacher said tickets in PDF form were “a nightmare” to sort.
- Teachers wanted access to current and previous KELPA scores in Educator Portal.

---

<sup>9</sup> Only open-ended questions are included in Appendix C. Therefore, question numbers match their order in the original survey and may not be consecutive.

“Access to a list of students that indicates which tests are completed, which [tests] are in progress, and which tests have not been started.”

“Being able to see my transfer [students’] scores if they took the KELPA in 2022 at another building, I’d like to be able to get their score sheet to print off along with the students that took the KELPA at my school during the testing window.”

“I would like to see what score is listed for each student’s test items as the BTC [Building Test Coordinator]. This would help to quickly identify who needs to have their score edited for a specific test # item. This year we had to completely redo 2 whole items for 5th graders because the teacher who scored couldn’t remember which students she scored a 0 due to no audio. As the BTC I cannot see the score a student gets. When I click on an item to edit, the score that was given is no longer there. I could hear all [students’ audio] just fine so I had to go through all of them so students didn’t get zeros. I want to be able to easily identify scores in these incidences, so I don’t have to do double the work.”

**Q6.** *If you could change one thing about Educator Portal, what would you change?*

- Teachers wanted to see a display of scores for each test item after scoring that is available to teachers, test coordinators, and building staff who support scoring.
- Teachers wanted to see a search box or tab to look up individual students, as well as a central location to access all student usernames and passwords for the Kite system.
- Teachers wanted a portal that is easier to navigate (including user-friendly tabs for important functions such as access codes, reactivation of tests, tests that need to be administered, etc.), with the default number of displayed items per page set to 30.
- Teachers wanted to see descriptions of expectations in each category.
- To inform instructional decisions, teachers wanted to see a detailed breakdown of skills students need to work on.
- Teachers wanted to see a view of multiple students on the same page.
- Teachers wanted a field in which to write and store notes.
- One respondent wanted to see a log of when password reminders are emailed to teachers so there is a record of password-change notification emails to assist with account management issues.

“I wish I could see scores I’ve previously given and have the option to change them. The rubrics are so terrible that often times after scoring other students I would realize I should have given a previously scored student a different score on an answer.”

“I would like to see the division of speaking and writing tests (for scoring) in a way that could be readily ‘marked’ and highlighted (along with the number) as scored or not scored.”

“Assigning additional scorers or Second-Rater scoring assignments for KELPA. It is a multi-step process and not easy to navigate. Because we have a small enough population, many of our students are at various buildings and/or grade levels. This meant our DTC [District Test Coordinator] had to edit and revise teacher access. This year, our DTC ended up adding multiple (of the same student) to teacher rosters. Maybe if the required fields could be edited or NOT required, multiple students can be selected and added to a teacher’s roster.”

**Q7. Is there anything we should change in Educator Portal to make scoring easier?**

- Teachers wanted less clicking and scrolling in the portal and want to view prompts and answers on the same page.
- Teachers found it frustrating that setting changes (such as number of students displayed at one time, paired ratings, etc.) are not retained throughout the session.
- Teachers wanted to see a list of information that is uploaded when performing a batch upload.
- Teachers wanted to see more detail in the rubrics that they use.
- Teachers did not want to select the type of scoring for each new question.
- Teachers described the scoring process as extremely time consuming, noting too many drop-down boxes to fill in, and they disliked continuous scrolling for each question and student.

“When scoring multiple students on multiple items, the flow is interrupted either when you get to the end of the student list or the end of the question list. It would be nice to continue scoring without being kicked back to the main page.”

“Entering the Speaking and Writing parts individually takes a long time. It takes 15–20 minutes to record the speaking part and then you have to get on Educator Portal and type in the scores—it is 3 extra ‘clicks’ per student the way the scoring is set up. That adds up to extra time per student! It would be better if you could click the rubric and then click next questions.”

“Personally, I would rather have the ability to click on more than JUST the one score at the top of the column. For instance, in ‘Addressing the topic’ there is more than one item to evaluate and often, the student fulfills one item at a different proficiency level. It would be nice to be able to click 3,2,1 on individual items within the rubric.”

“Let there be more than 10 students that can be viewed at one time without having to toggle the drop-down arrow every time you switch grade bands, freeze the top row so that no matter how far you scroll down you always can see what question number you are on, show everything needed for grading on one screen with no scrolling up or down and not opening a separate tab to see the stimulus, and change the rubrics (clumping...performing the task with their language abilities is not the best way to score an ELL test).”

#### **IV. Scoring**

**Q5. Please list materials your district used for rater training.**

Respondents named the following materials:

- rubrics (writing and speaking)
- sample responses
- *KELPA Examiner Manual*
- *KELPA Manual*
- *Rater-Training Manual*
- practice test guide
- materials from KSDE site
- materials from Kite Portal
- PowerPoint slides
- audio files
- webinars/training videos

- test security and ethics training
- information from district ESOL website

Some teachers were simply emailed documents and told to read them; others were directed to visit the Kite site and obtain all rater training on their own, without any guidance from their district.

## VI. Utility of KELPA

**Q6.** *If KELPA test results are or will be used for purposes other than informing future ESOL services, please indicate and/or describe.*

Respondents described the following uses for results:

- to determine the number of minutes needed for services
- to determine the need for special education services
- to determine materials and modifications for in-class assignments
- to determine whether a student should be enrolled in Placement of Newcomers program
- to determine student placement in English language arts and mathematics classes
- to determine whether students qualify for the Seal of Biliteracy
- to differentiate instruction for students
- when using the problem-solving process for a student
- to form intervention groups

“I wish there were other ways for a student to exit ESOL support as some ELs (who are orally proficient) don’t perform to their potential on the speaking domain.”

“Students who qualify to take the KELPA at my building have waived their ESOL services. The results will be shared with grade-level teachers as another piece of data to help guide/drive instruction.”

“We try to use the KELPA results to align to ELP [English language proficiency] standards and give more specific strategies to classroom teachers. These are typically outlined on the Individual Learning Plan (ILP). We like to look at KELPA data and triangulate the data between other assessments and screeners—most specifically looking at our long-term ELs. I believe KELPA is (or can be) used for Seal of Biliteracy, but I’m not sure if we have many ELs moving toward that goal—or even aware of it. Why can’t KELPA (computerized) test directions be translated and available for students to listen too? The directions are essentially the same on each test and each grade band. If students (especially our newcomers) just had a better understanding of the assessment and [its] purpose, I think there would be a difference in outcome.”

“I am able to see growth, but not areas of concern.”

“The scores help with nothing! We get no feedback at all!”