



# **Test Administration Technical Report**

2023–2024 Program Year

**February 2025**

**Prepared for:**  
**Ohio State Board of Education**  
**William Green Building**  
**30 West Spring Street, 12th Floor**  
**Columbus, Ohio 43215**

THIS PAGE LEFT INTENTIONALLY BLANK

## Table of Contents

<b>Preface.....</b>	<b>4</b>
<b>Purpose of the Ohio Assessments for Educators (OAE) Program .....</b>	<b>5</b>
Introduction .....	5
Composition of the OAE Program .....	5
<b>Technical Properties of the OAE Program.....</b>	<b>7</b>
Scoring .....	8
Item Analyses .....	10
Test Equating .....	11
Scaled Scores.....	11
Reliability .....	14
Validity.....	17
<b>Score Reporting .....</b>	<b>19</b>
Candidate Score Reports.....	19
Other Score Recipients .....	20
Test Statistics Reports .....	20
Pass Rate Summary .....	22
Title II Reporting .....	23
Support for Veterans: Reimbursements for Educator Licensure and Testing .....	23
<b>References .....</b>	<b>24</b>
<b>Appendices .....</b>	<b>25</b>
Appendix A: Sample OAE Candidate Score Reports: Pass and Did Not Pass .....	26
Appendix B: Test Form Statistics Report for Test Fields with More than 100 Tests Taken .....	35
Appendix C: Test Form Statistics Report for Test Fields with Fewer than 100 Tests Taken .....	53
Appendix D: Test Field Statistics Report for Test Fields with Constructed-Response Items.....	57
Appendix E: Pass Rate Summary.....	60

## Preface

---

This test administration technical report provides information on the technical characteristics of the tests and score reporting for the Ohio Assessments for Educators (OAE) program for the 2023–2024 program year from September 1, 2023, through August 31, 2024.



# Purpose of the Ohio Assessments for Educators (OAE) Program

---

## Introduction

The Ohio Assessments for Educators (OAE) program, administered by Pearson, assesses the content-area and professional (pedagogical) knowledge of candidates who are either seeking initial Ohio educator licensure or adding a new licensure area. The OAE tests are aligned with Ohio Educator Standards, Ohio Learning Standards, and other professional standards, as appropriate.

Each test was validated for use in Ohio in accordance with the practices recommended by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). The Standards require a clear definition of content domain and a rationale to support a claim that the knowledge, skills, and abilities being assessed in a licensure test are required for credential-worthy performance. Educators, educator preparation program faculty, and administrators from across Ohio were involved in reviewing the test materials for content, job-relatedness, and fairness; validating their appropriateness for use in Ohio; and making recommendations for the passing score for each test.

The OAE tests are computer-based and delivered through a national network of Pearson computer-based testing centers. Most tests are available year-round by appointment. Currently 30 tests are available via online proctoring in addition to test center-based administration.

The OAE program offers several web-based resources to help candidates prepare for the tests. These resources include online study guides, practice tests, detailed score reports, and computer-based testing tutorials. In addition, a suite of faculty resources and interactive worksheets are available to assist in candidate preparation. The Ohio State Board of Education (SBOE) and educator preparation programs have access to an interactive, electronic database that allows them to create customized reports of candidate test results and institution performance, or to perform customized data queries.

## Composition of the OAE Program

During the 2023-2024 program year, 57 OAE tests were available for test administration, including 45 content-area assessments<sup>1</sup> and five professional (pedagogical) knowledge assessments. Professional (pedagogy) knowledge tests are matched to Ohio licensure grade bands—Early Childhood, Primary (PK-5), Middle Childhood (4-9), Adolescence to Young Adult (7-12), and Multi-Age (PK-12). Content-area tests match Ohio license types.

The following new assessments were first administered in the 2023-2024 program year: the Dual Adolescence to Young Adult Special Education (7-12), and Dual Middle Childhood Special Education (4-9) on September 18, 2023; Financial Literacy on February 5, 2024. The Ohio Assessments for Educators (OAE) test titles are provided on the next page.

---

<sup>1</sup> Six OAE content-area assessments include two separate tests each (i.e., Subtest I and Subtest II) for a total of 51 unique tests.

## **Ohio Assessments for Educators (OAE)**

### **Pedagogical Knowledge Assessments:**

- 001 Assessment of Professional Knowledge: Early Childhood (PK–3)
- 057 Assessment of Professional Knowledge: Primary Education (PK–5)
- 002 Assessment of Professional Knowledge: Middle Childhood (4–9)
- 003 Assessment of Professional Knowledge: Adolescence to Young Adult (7–12)
- 004 Assessment of Professional Knowledge: Multi-Age (PK–12)

### **Content Knowledge Assessments:**

- 005 Agriscience
- 050 American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest I)
- 051 American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest II)
- 006 Art
- 007 Biology
- 008 Business Education
- 009 Chemistry
- 054 Computer Science
- 016 Computer/Technology (Subtest I)
- 017 Computer/Technology (Subtest II)
- 011 Dance
- 061 Dual Adolescence to Young Adult Special Education (7–12)
- 060 Dual Middle Childhood Special Education (4–9)
- 012 Early Childhood Education
- 013 Early Childhood Special Education
- 014 Earth and Space Science
- 015 Educational Leadership
- 018 Elementary Education (Subtest I)
- 019 Elementary Education (Subtest II)
- 020 English Language Arts
- 021 English to Speakers of Other Languages
- 022 Family and Consumer Sciences
- 059 Financial Literacy
- 190 Foundations of Reading (FOR)
- 053 Gifted Education
- 023 Health
- 024 Integrated Science
- 025 Integrated Social Studies
- 026 Marketing
- 027 Mathematics

## **Ohio Assessments for Educators (OAE) (continued)**

028 Middle Grades English Language Arts  
029 Middle Grades Science  
030 Middle Grades Mathematics  
031 Middle Grades Social Studies  
032 Music  
034 Physical Education  
035 Physics  
036 Prekindergarten (Subtest I)  
037 Prekindergarten (Subtest II)  
038 Reading (Subtest I)  
039 Reading (Subtest II)  
040 School Counselor  
041 School Library Media Specialist  
042 School Psychologist  
043 Special Education  
044 Special Education Specialist: Deaf/Hard of Hearing  
045 Special Education Specialist: Visually Impaired  
046 Technology Education (Subtest I)  
047 Technology Education (Subtest II)  
048 Theater

## **Technical Properties of the OAE Program**

---

The *Standards for Educational and Psychological Testing* requires that testing agencies provide relevant technical information about the assessments so that test users and reviewers have sufficient information to make judgments about the quality of the test, the resulting scores, and interpretations based on test scores (AERA, APA, & NCME, 2014). This information can ultimately assist test users and reviewers in determining the appropriateness of the test for its intended purpose (AERA, APA, & NCME, 2014).

## Scoring

The OAE program includes tests consisting of multiple-choice items only as well as tests consisting of both multiple-choice and constructed-response items. The scoring procedures for the program are carefully documented for both multiple-choice and constructed-response items. Additionally, performance monitoring is conducted to check the accuracy of scoring and reporting for the constructed-response items.

### Scoring Multiple-Choice Items

Answer keys for multiple-choice items are prepared during the construction of test forms. These keys are reviewed and checked at several points during development and production. Automated technology at computer-based testing centers compares these answer keys to candidate responses immediately following the conclusion of testing. This technology allows for on-site unofficial test results to be generated at the testing centers, which can then be provided to candidates. These unofficial test results are provided to candidates only for the OAE assessments that do not include constructed-response items.

Multiple-choice items are dichotomously scored, meaning a single point is awarded for each correct response, and no points are awarded for an incorrect response. For tests composed of multiple-choice items only, the final raw score is the total number of correct responses on the test. The raw scores are transformed and reported on a scale ranging from 100 to 300 with a scaled passing score of 220.

Each test form for all OAE fields includes both scorable and nonscorable multiple-choice items. Scorable items are those that are used to compute candidates' scores. Nonscorable items are those that are included on a test form to collect additional psychometric information (to support pilot testing) and to support test form equating methodologies but do not contribute to candidates' scores.

### Scoring Constructed-Response Items

Some OAE tests include constructed-response items in addition to multiple-choice items. Candidate responses to constructed-response items are scored using a focused holistic scoring methodology. In this method, scorers judge the overall effectiveness of each response using a set of characteristics that have been defined as important to help inform the overall score. The score is holistic in that each score is based on the overall effectiveness of these characteristics working together, focusing on the response as a whole.

The *Special Education Specialist: Visually Impaired* constructed-response item is scored differently. This item requires candidates to transcribe a passage into Braille, and the score provided is based on the number of errors made in the transcription.

*American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest II)*: Candidates are video-recorded as they respond to four constructed-response items. For tests taken prior to October 16, 2023, a participatory scoring model was used, whereby candidates were expected to primarily use ASL to respond to the constructed-response items and responses were reviewed by Pearson's scoring team to confirm scorability.

A passing standard for candidate responses, recommended by a standard setting panel convened in December 2022, was applied beginning with tests administered October 16, 2023.

**Scoring components.** Candidate responses are scored on a scale. Some tests use a four-point score scale; other tests use a three-point score scale. A score of 1 represents little or no command of the characteristics, and 3 (for those on a three-point scale) or 4 (for those on a four-point scale) represents a strong command of the characteristics. Each constructed response is independently scored by two scorers, and these scores are summed for a total possible score range of 2 to 6 for three-point score scales or 2 to 8 for four-point score scales. If the two scores are not identical or adjacent, additional scoring is conducted to resolve the discrepancy.

**Scorer qualifications.** To participate in the holistic scoring process, scorers must meet specific qualifications such as:

- A state educator certificate/license;
- Experience as an educator in public schools; and/or
- Experience as a college faculty member responsible for preparing prospective educators.

**Scorer training.** Prior to each scoring session, scorers receive training and orientation by a Chief Reader who trains each scorer to evaluate responses to a variety of educator licensure examinations. The Chief Reader provides each scorer with the background of the OAE tests, the context of the scoring task, the tasks they will perform, the procedures they will follow, the scoring scale, and the characteristics that will guide their scoring judgments. The scorers are oriented to the prompt-specific rubrics that apply the characteristics and scoring scale to the constructed-response item(s) used on test forms for the current test administration. The Chief Reader identifies and develops orientation materials, leads scoring sessions, conducts calibration orientation, and monitors the progress of the scoring session.

Scorers receive practice using training sets of responses to which scores have already been assigned, including marker responses (i.e., exemplar responses of each score point on the score scale). The training emphasizes consistent application of the score scale. Once scorers have been trained, they have to demonstrate their ability to score accurately by completing a calibration exercise.

**Performance Monitoring of Scorers.** Pearson monitors the performance of scorers throughout the scoring sessions and implements checkpoints with scorers. Specific areas monitored include a scorers' ability to understand and apply the established scoring scale, the consistency of the scores assigned in comparison with those assigned by the second scorer, and the scorers' consistency over time. During checkpoints, scorers are recalibrated to the scale, typically through discussions of specific questions raised by scorers. Scorers must demonstrate continued scoring accuracy on the responses. If scorers fail to demonstrate accuracy, they receive additional, individual reorientation before proceeding with scoring.

## Item Analyses

Item analyses are conducted on multiple-choice items to assess the accuracy and psychometric quality of the items. Additionally, data from constructed-response items are reviewed to confirm that items in the item bank for each field are comparable in terms of difficulty and score distribution.

### Item Analysis for Multiple-Choice Items

The purpose of item analysis for multiple-choice items is to verify the accuracy of the answer key for each administered test form and as an additional quality assurance check before providing final results and before official candidate score reports are produced.

Data are collected on each item, allowing for the empirical consideration of item difficulty, item discrimination, content accuracy, and the plausibility of distractors. These item statistics are calculated and evaluated for the current administration and cumulatively (i.e., using combined statistics from previous operational test administrations).

The item statistics calculated and evaluated for each multiple-choice item include:

- Item difficulty (p-value);
- Distribution of responses (percentages of participants selecting each response option);
- Item-to-total test point biserial correlation (correlation of performance on the item to performance on the total test); and
- Mean score by response choice (average score on the total multiple-choice set achieved by all participants selecting each response option).

Those items that do not perform within defined statistical parameters are flagged and reviewed. Flagged items are reviewed by content specialists, test development specialists, psychometricians, and editors. Reviewed items are deleted or revised and subjected to additional pilot testing. The review confirms that the wording on the test forms matches the wording validated by the content advisory committees. During the review, there is a check of content, topicality, and correct answer.

### Item Review for Constructed-Response Items

Constructed-response item data are regularly reviewed to monitor the difficulty of the items across administrations. Throughout the scoring process, monitor reports containing mean scores and standard deviations are reviewed. The monitor reports also contain distribution of scores assigned by the first two scorers, distribution of score differences, and distribution of score combinations assigned by the first two scorers.

## Test Equating

Each OAE test consists of multiple forms. Multiple forms are utilized within and across test administrations to address potential item exposure and maintain security. Industry-accepted statistical adjustments (equating) are implemented to adjust for small differences in difficulty across forms.

### Purpose of Equating

According to the *Standards for Educational and Psychological Testing*, equating refers to the process of placing scores from alternate, parallel, or equivalent forms of a test on a common scale (AERA, APA, & NCME, 2014). The central purpose of statistical equating methods is to compensate statistically for possible variability in the characteristics of test forms that may affect candidates' scores (e.g., differences in the overall difficulty of a new test form compared to a previous test form). Statistical equating methods ensure that a candidate's scaled score is adjusted for the relative difficulty of the particular test form that was taken. Equating thus allows test developers to attribute differences in scores across test forms to differences in knowledge or skills, and not differences in the tests. Equating helps to ensure that different test forms are comparable and that test scores from multiple test forms can be used interchangeably.

### Equating Design

A single-group equating design is utilized for the OAE tests. In a single-group design, the same group of candidates is scored on two alternative forms of the same test. The two forms are defined by designating two alternative subsets of items on the full test form as scorable, so that each alternative contains a set of common scorable items plus a set of scorable items unique to the form. The two alternative forms are then statistically equated.

To equate the two test forms, a linear equating method was implemented for the OAE program. In linear equating, two scores are equivalent if they are the same number of standard deviation units above or below the mean for some group of candidates (Angoff, 1984). A linear equation is used to relate the scores from the two forms by setting standard deviation scores, or z-scores, to be equal on the two test forms (Kolen & Brennan, 2004).

This design is used for the OAE tests because of the need to pre-equate the test forms, enabling candidates to receive unofficial test results at the testing center immediately after the conclusion of their test administration (for multiple-choice-only OAE tests). With pre-equating, the passing score for a new test form is established prior to operational administration.

## Scaled Scores

The *Standards for Educational and Psychological Testing* states that scaled scores may aid in interpretation of the test. Specifically, scaled scores allow scores to be easily comparable regardless of test form or administration (AERA, APA, & NCME, 2014).

Scaled score reporting is preferred to raw score reporting due to the confusion that may occur as a result of some changes in raw cut scores across test forms. Therefore, the use of scaled scores helps to support the communication of the OAE program results in the following ways:

- Candidates, educator preparation programs, and stakeholders will be able to interpret scores from the different OAE tests in a similar manner, regardless of the test taken.
- The meaning of the scaled passing scores will be consistent over time, making it possible to compare performance from one test administration to the next.

### Scaled Score Range for the OAE

Raw test scores, including the number correct for multiple-choice items and holistic scores for constructed-response items, are transformed to a standard scale ranging from 100 to 300. This scale is used to communicate OAE scores across all tests within the program. For each OAE test, 220 represents the passing score. By using a standard scale across all tests, test scores and feedback are provided in an easy-to-understand format for candidates, educator preparation programs, the SBOE, and other stakeholders.

### Calculation of Scaled Scores

For tests containing only multiple-choice items, a simple linear transformation is applied to the raw scores to compute the associated scaled score. The scaled score is derived from the candidate's raw score, the raw cut score, and the maximum possible raw score. Candidates who perform at the raw score that is equivalent to the passing score achieve a scaled score of 220, while those who achieve the maximum possible raw score will receive a scaled score of 300.

For the OAE tests that contain both multiple-choice and constructed-response items, the multiple-choice and constructed-response test sections are each scaled separately. These multiple-choice and constructed-response scaled scores are then combined in accordance with the following weightings.

OAE Test	Multiple-Choice Weight	Constructed-Response Weight
Assessment of Professional Knowledge: Early Childhood (PK-3)	80%	20%
Assessment of Professional Knowledge: Primary Education (PK-5)	80%	20%
Assessment of Professional Knowledge: Middle Childhood (4-9)	80%	20%
Assessment of Professional Knowledge: Adolescence to Young Adult (7-12)	80%	20%
Assessment of Professional Knowledge: Multi-Age (PK-12)	80%	20%
American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest II)	40%	60%
Computer/Technology (Subtest I)	70%	30%
Computer/Technology (Subtest II)	70%	30%
Educational Leadership	60%	40%
Foundations of Reading (FOR)	80%	20%
Prekindergarten (Subtest I)	80%	20%
Prekindergarten (Subtest II)	80%	20%
Reading (Subtest I)	80%	20%
Reading (Subtest II)	80%	20%
School Psychologist	85%	15%



<b>OAE Test</b>	<b>Multiple-Choice Weight</b>	<b>Constructed-Response Weight</b>
Special Education: Visually Impaired	80%	20%
Technology Education (Subtest I)	70%	30%
Technology Education (Subtest II)	70%	30%
Theater	80%	20%

## Reliability

The *Standards for Educational and Psychological Testing* refers to reliability as the consistency of test scores for a group of candidates across administrations (AERA, APA, & NCME, 2014). There are many common reasons for individual scores to fluctuate over time. Score fluctuation from one testing occasion to another has an impact on reliability. Some factors that affect reliability include:

- **Number of candidates.** The number of candidates whose test scores contribute to a statistical estimate of reliability affects the stability of the estimate. In general, reliability estimates based on larger numbers of candidates are more stable than estimates based on smaller numbers. For this reason, reliability estimates are calculated for tests that are taken by 100 or more candidates.
- **Self-selection of candidates by test administration date.** Typically, candidates can decide when to take a particular test. OAE tests are administered throughout the year, and candidates can select when to take and retake the tests. This self-selection can affect the composition, ability level, and variability of the group taking a particular test at a given test administration.
- **Variability of the group tested.** In general, the larger the true variance or true spread of the scores of the candidate group (i.e., the greater the individual differences in the true level of knowledge, skills, and abilities of the candidates in the particular group taking a test on a particular occasion), the greater will be the reliability coefficient. Reliability estimates tend to be higher if candidates in the group have widely varying levels of knowledge, and lower if they tend to have similar levels of knowledge.
- **Test length.** Statistical estimates of reliability are typically higher for tests with greater numbers of questions. A more reliable estimate of a person's knowledge is obtained by asking more questions.
- **Test content.** Reliability estimates are typically higher for tests that cover narrow, homogeneous content than for tests (such as many used for educator licensure) that cover a broad range of content. Tests for educator licensure must typically test a broad base of knowledge, skills, and abilities that pertain to licenses that will apply in a wide range of educational settings, grade levels, and teaching assignments.

Because the tests included in the OAE program are used to make high-stakes decisions, several indicators of decision consistency (that is the degree to which the same decisions are made from two tests) and measures that indicate score reliability (consistency of scores across administrations) are calculated. Statistics presented not only consider the reliability of the test scores, but also indicate the reliability of the decisions made using the test results.

Several measures are employed to assess the reliability of each test in the OAE program. These measures are described below.

- **Livingston-Lewis estimate of decision consistency.** For a test used to make licensure requirement decisions such as those in the OAE program, the consistency of such decisions becomes a primary focus (Crocker & Algina, 1986). Decision consistency refers to the degree to which the same decisions are made from two tests. For the OAE program, the Livingston and Lewis (1995) estimate of decision consistency is used. This multi-stage method calculates decision consistency and accuracy using four types of input, including distribution of scores on one form, the minimum and maximum possible scores, the cut points used for classification, and the reliability coefficient (Livingston & Lewis, 1995). Decision consistency is reported in the range of 0 to 1, with estimates close to 1 indicating more consistent or reliable decisions.
- **Kuder-Richardson formula 20 (KR20).** The Kuder-Richardson index of item homogeneity (KR20) is an overall test consistency (reliability) estimate based on a single test administration (Kuder & Richardson, 1937). It is applicable to the multiple-choice section of tests. KR20 is reported in the range 0 to 1, with a higher number indicating a greater level of consistency (reliability). Homogeneity refers to the degree to which the items on the test are consistent with one another. For the OAE program, KR20 is computed for tests composed of multiple-choice items only as well as for multiple-choice sections of tests that also include constructed-response items.
- **Stratified coefficient alpha.** Stratified coefficient alpha is an estimate of total test score reliability for a test containing a mixture of item types (e.g., both multiple-choice and constructed-response) (Qualls, 1995). Each item type component of the test is treated as a subtest. Internal consistency estimates for the separate subtests are combined to compute stratified coefficient alpha. Stratified coefficient alpha is reported in the range 0 to 1, with a higher number indicating a greater level of consistency (reliability). This statistical estimate was deemed most appropriate for estimating total reliability of tests with both multiple-choice and constructed-response items for the OAE program because it takes into account differences in test length and variance of the two item types.
- **Standard error of measurement.** The *Standards for Educational and Psychological Testing* defines the standard error of measurement as the estimate of the difference between observed scores and estimated true scores by estimating the variability of measurement errors. This statistic speaks to the reliability of test scores, with smaller standard errors of measurement indicating more reliable test scores (AERA, APA, & NCME, 2014).
- **Generalizability coefficient (G).** The Generalizability (G) coefficient is a measure of the percent of total score variance that is attributable to persons (i.e., factors within the candidate, such as subject matter knowledge). It reflects the proportion of variability in individuals' scores that is attributable to true score variability rather than to measurement error (Brennan, 2001). It is reported in the range 0 to 1, with a higher number indicating a greater level of generalizability. The G-coefficient is applicable to test sections composed of constructed-response items. It gauges the degree to which the results from one test form of the constructed-response items are generalizable to other forms or other test administrations. The G coefficient for exam forms with a single constructed-response item is calculated with the item type as a random facet. This permits variability of scores to be associated with the item type rather than

with error variance (Brennan, 2001). Some exam forms have more than one type of constructed response item. Calculating G coefficient for these exam forms as a random facet will result in lower G coefficients.

- **Scorer agreement.** Scorer agreement is the degree of agreement between constructed-response scores assigned by independent scorers. Independent scorers are in agreement if the scores they award are either exact or adjacent. The scorers are not in agreement if the scores awarded differ by more than one point. The percent of cases in which the first two independent scorers are in agreement is computed as a measure of scorer agreement (reliability). The following scorer agreement statistics are reported.
  - *Percent agreement.* Overall agreement determined by summing exact and adjacent agreement.
  - *Percent exact.* Percentage of scores in which the first two scorers were in exact agreement.
  - *Percentage adjacent.* Percentage of scores in which the two scorers assigned adjacent scores.
  - *Inter-rater reliability.* Intraclass correlation between the first and second score assigned to each response, corrected using the Spearman-Brown formula.

## Validity

The *Standards for Educational and Psychological Testing* states that validity is a fundamental consideration in developing and evaluating tests (AERA, APA, & NCME, 2014). Validity relates to the use and interpretation of test scores rather than describing a test itself. For the OAE program, collection of both content-based and construct-based validity evidence is a continuous process.

### Content-Based Validity Evidence

Because the OAE program is composed of licensure tests, gathering content-based validity evidence is essential to confirm that the assessment frameworks (which represent the content domain) are representative of the knowledge, skills, and abilities necessary for an entry-level educator in Ohio. Content-based validity evidence for the OAE program was gathered throughout the assessment review and validation process, starting with the frameworks.

**OAE frameworks.** In reviewing, revising and validating the frameworks, content-based validity evidence is initially gathered through an alignment study. Alignment can be used to gather content-based validity evidence by corroborating that the knowledge, skills, and other constructs measured by the tests are consistent with those specified in the frameworks (Koretz & Hamilton, 2006). The OAE assessment frameworks are based on state-approved and nationally recognized professional and academic standards and contain a competency component with information about the knowledge and/or skills necessary for performing the job of a licensed educator in Ohio public and non-public schools. Therefore, the competencies collectively define the range of content to be measured by the assessment. Pearson conducted an alignment study of each OAE framework. These alignment studies were carried out to confirm that the test content, and therefore the tests, are aligned with appropriate standards related to the intended purpose of the test. Additional information on the alignment studies can be found in the Development and Validation Technical Reports on the Ohio Assessments for Educators program website. The [OAE Technical Reports](#) are available under Faculty Resources on the Ohio Assessments for Educators website.

Another source of content-based validity evidence was gathered by conducting an initial review of the draft frameworks by Ohio educators and Ohio educator preparation program faculty as part of Fairness and Content Advisory Committees. During these reviews, Ohio educators and educator preparation program faculty checked the content domain represented the knowledge, skills, and abilities required by an entry-level educator in that test field by considering the criteria of alignment, completeness, clarity of language and terminology, and fairness.

**OAE content validation surveys.** Content validation surveys provide content-based validity evidence through the input of experts and stakeholders in educator licensure regarding the importance of the necessary knowledge, skills, and abilities specified in each framework for an entry-level educator. Results of these surveys guided the final definition of the content domain. Ohio educators and Ohio educator preparation program faculty rated the test components of each framework on a 1–5 scale. Results of the surveys were used to determine the final status of all the assessment framework components. Results of the content validation surveys for the OAE indicated that across test fields, the majority of competencies and descriptive statements achieved a rating of at least 4.0, representing “great importance.” Additional information on the content validation surveys is included in the Development and Validation Technical Reports on the Ohio Assessments for Educators program website. The [OAE Technical Reports](#) are available under Faculty Resources on the Ohio Assessments for Educators website.

**OAE draft test items.** As described in the Development and Validation Technical Reports, the item review and validation activities by the Fairness Advisory Committee and Content Advisory Committees provided additional content-based validity evidence. The Fairness Advisory Committee made recommendations to the content advisory committees to revise the draft items for accessibility and fairness and to ensure items represented the diversity of the Ohio population. The Content Advisory Committees reviewed and revised the draft items as necessary, then validated all items to indicate that they matched the test competency or content domain to which they were written, and were accurate, fair, and job-related.

### Construct-Based Validity Evidence

The *Standards for Educational and Psychological Testing* refers to construct validity as the degree to which scores from an assessment can be interpreted as indicating the candidate's standing on the knowledge, skills, and abilities assessed by the test (AERA, APA, & NCME, 2014). Some threats to construct validity include construct irrelevance and construct underrepresentation. Construct irrelevance is "variance in test-taker scores that is attributable to extraneous factors that distort the meaning of the scores and thereby decrease the validity of the proposed interpretation" (AERA, APA, & NCME, 2014, p. 217). Construct underrepresentation is defined as "the extent to which a test fails to capture important aspects of the construct domain that the test is intended to measure" (AERA, APA, & NCME, 2014, p. 217).

**OAE content validation surveys.** For the OAE program, Pearson conducted content validation surveys that were aimed at gathering evidence to show that each assessment framework adequately reflects the knowledge, skills, and abilities necessary for an entry-level educator in Ohio. Survey participants were asked to rate the importance of the knowledge, skills, and abilities described by each competency and accompanying descriptive statements. The third item on the survey asked participants to rate a set of competencies in terms of how well they represented important aspects of the knowledge, skills, and abilities required for performing the job of an entry-level educator. Results showed that the set of competencies for each test adequately represents the knowledge, skills, and abilities the test is intended to measure.

**OAE test items.** The item review and validation activities by the Fairness and Content Advisory Committees provided additional construct-based validity evidence. The Fairness Advisory Committee made recommendations to the content advisory committees to revise the draft items for accessibility and fairness and to ensure items represented the diversity of the Ohio population. The Content Advisory Committees reviewed and revised the draft items as necessary, then validated all items to indicate that they matched the test competency or content domain to which they were written, and were accurate, fair, and job-related.

## Score Reporting

---

After administration of the OAE tests, official score reports are provided to candidates to inform them of their passing status and performance on the test. Official score reports are also provided to SBOE and to Ohio educator preparation programs, as designated by the candidate, to communicate passing status of individual candidates for educator licensure.

### Candidate Score Reports

Candidates receive unofficial test results (reported as pass or not pass) on-site following the administration of multiple-choice-only tests. For tests that contain constructed-response items, candidates receive a receipt of completion on-site following the test administration. All candidates may request that an official score report be e-mailed to the address provided during the registration process on the score report date published on the program website. Official score reports are posted to the candidate's online account for two years as PDF documents, which a candidate may view, print, and save for their records. After the two-year period, candidates may request a copy of their scores through their online account.

Official score reports include the following information.

- the date the candidate took the test
- the candidate's overall scaled score based on the number of items answered correctly converted to a scale ranging from 100 to 300
- the candidate's passing status based on the state-approved passing standard
- details of candidate performance on each content domain assessed by the test for both multiple-choice and constructed-response sections

Candidates who do not pass a test are also provided with a Detailed Performance Summary that includes specific information about the candidate's performance on each test competency and constructed-response item, if applicable.

Sample *Pass* and *Not Pass* candidate score reports are provided in Appendix A. Score reports are accompanied by an interpretive guide to help candidates understand the reports. Additional information on [how to read the score report](#) can be found on the OAE program website.

## Other Score Recipients

Candidate test results are sent to the State Board of Education (SBOE) and educator preparation programs as designated by the candidate. These scores are delivered electronically through Pearson's secure web-based score-reporting system. Through this system, the SBOE and educator preparation programs have access to a web-based tool called *ResultsAnalyzer*®. This tool is interactive and allows the SBOE and educator preparation programs to view, analyze, reorganize, download, and print results based on test data and generate customized reports of their choice.

*ResultsAnalyzer*® can support the needs of Ohio educator preparation programs through the following features.

- create relevant and timely reports on candidate results
- filter reports by candidate demographics, specific tests, or testing year
- organize data into customized tables and graphs
- customize data queries to align with institution goals and areas of interest
- aggregate performance data across testing program years
- export data to Microsoft® Excel or other report software and print graphics
- analyze candidate data for numerous purposes, including Title II reporting, educator preparation program research, program evaluation, and curriculum improvement

*ResultsAnalyzer*® also allows the SBOE to generate reports based on its needs and areas of focus using the features previously mentioned. In addition, *ResultsAnalyzer*® can create reports and filter results by educator preparation program, as well as support the analysis of institutional data. These data can assist the SBOE in assessing the performance of educator preparation programs across the state and help in future statewide policy decisions.

## Test Statistics Reports

This document provides statistical reports for each test and reports on candidate performance for OAE tests administered during the 2023–2024 Program Year. Test statistics reports consist of test form and field statistics reports that outline statistical characteristics of multiple-choice-only tests as well as tests composed of both multiple-choice and constructed-response items. Total scaled score distributions are provided. Specific details of each report are outlined below.

### Test Form Statistics Reports

The Test Form Statistics Report in Appendix B provides selected statistics for test forms administered to at least 100 candidates during the program year. This report includes the following information:

- *Test Field*. Code identifier and name assigned to each test.
- *Form*. Identifier assigned to each form.
- *Number of tests taken*. Scores are included in the analyses for this report for candidates who attempted at least one multiple-choice item and provided scorable responses to the constructed-response items, if applicable. In the cases



in which the same test forms were administered at more than one test administration during the reporting period and candidates took the same test form at more than one test administration, both scores for those candidates' several attempts are included in the analyses.

- *Mean*. The average of the total test scaled scores achieved by the candidates taking the test form.
- *Standard error of measurement (SEM)*. The SEM of the total test scaled scores is an estimate of the variability of measurement errors by approximating the difference between observed scores and estimated true scores.
- *Decision consistency*. The Livingston-Lewis estimate of decision consistency is based on all administrations of a test form during the program year. Test forms are considered to be identical if they contain identical sets of scorable multiple-choice items, regardless of the order of the items.
- *Stratified alpha*. This statistic provides an estimate of total test score reliability for a test containing multiple item types (e.g., both multiple-choice and constructed-response items) (Qualls, 1995).
- *Test length*. The number of scorable items of the multiple-choice section or constructed-response sections.
- *KR20*. The Kuder-Richardson index of reliability applicable to the multiple-choice sections.
- *G coefficient*. Gauges the degree to which the results from one test form of the constructed-response items are generalizable to other forms or administrations.

A second Test Form Statistics Report is provided in Appendix C: *Test Form Statistics Report for Test Fields with Fewer than 100 Tests Taken*. This report contains the total test scaled score mean and the number of scorable multiple-choice items for test forms that were administered to fewer than 100 candidates during the 2023–2024 Program Year. Statistical indices such as the SEM or reliability coefficients cannot be confidently interpreted when computed using data from small groups of candidates; therefore, such analyses were omitted for test forms administered to fewer than 100 candidates in the reporting period.

## Test Field Statistics Report for Test Fields with Constructed-Response Items

The Test Statistics Report by Test Field shown in Appendix D provides selected statistics for test fields composed of constructed-response items administered to at least 100 candidates during the program year. This report includes the following information for each prompt/assignment.

- **Number of tests taken.** The number of times the test was taken during the program year.
  - *Number of valid scores.* The number of scorable responses that were included in the scorer agreement calculations.
- **Scorer agreement.** Measures of scorer agreement reported are:
  - *Percent agreement.* Overall agreement determined by summing exact and adjacent agreement.
  - *Percent exact.* This is the percentage of scores in which the first two scorers were in exact agreement.
  - *Percentage adjacent.* This is the percentage of scores in which the two scorers assigned adjacent scores.
  - *Inter-rater reliability.* This is the intraclass correlation between the first and second score assigned to each response, corrected using the Spearman-Brown formula.

The following fields are not included in Appendix D due to sample sizes below 100: 016/017 Computer/Technology Subtest I/II, 036/037 Prekindergarten Subtest I/II, 042 School Psychologist, 045 Special Education Specialist: Visually Impaired, 046/047 Technology Education Subtest I/II, 048 Theater, and 051 American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest II).

## Pass Rate Summary

The Pass Rate Summary shown in Appendix E provides data for examinees' initial (first time) and best attempts. The number of examinees and the percent at or above the passing score (if number of test takers is 10 or greater) are provided for each OAE assessment.

The pass rate summary includes:

- **N** = number of examinees who took the test.
- **%** = the percent of examinees at or above the passing score of 220.
- **Initial Attempt** = first occasion of an examinee taking the test. (2021-2022, 2022-2023, 2023-2024)
- **Best Attempt** = occasion on which an examinee achieved their highest test score. (2021-2024)

## Title II Reporting

The federal government requires that states submit annual reports on the quality of their educator preparation programs for accountability purposes. The data are intended to inform stakeholders on the quality of teacher preparation.

Pearson helps the Ohio Department of Higher Education (ODHE) meet Title II requirements by working with both the ODHE and educator preparation programs to produce the reports. Data exchange protocols with other testing companies that administer educator licensure exams adopted by the SBOE (such as the American Council on the Teaching of Foreign Languages, ACTFL) have been established for inclusion in the data collection and reporting process. The Title II reporting system allows the submission, review, and transmission of data that supports collaboration among educator preparation programs and the ODHE. Pearson supports the reporting needs and requirements through the following actions:

- producing Test Pass Rates Tables and Summary Pass Rates Tables for Title II Reporting.
- providing web-based training to educator preparation programs on the purpose and process for data collection.
- maintaining a secure, password-accessible website for data collection.

## Support for Veterans: Reimbursements for Educator Licensure and Testing

The Ohio Department of Education and Workforce has eliminated fees for educator licenses to recognize the contributions of military families. The fees for any initial Ohio educator license, permit or certificate – or for a renewal – will be waived for candidates who are veterans with honorable discharges or current service members of all branches of the United States Armed Forces, the National Guard or Reserve, or the Ohio Military Reserve or Ohio Naval Militia (under the Ohio Adjutant General). Spouses of active-duty service members also may receive a license free of charge.

The U.S. Department of Veterans Affairs offers a reimbursement program for educational expenses, which can include the cost of taking the Ohio Assessments for Educators. For more information, visit the [Department of Education and Workforce](#) website.

## References

---

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Angoff, W. H. (1984). *Scales, Norms and Equivalent Scores*. Princeton, NJ: Educational Testing Service.
- Brennan, R. L. (2001). *Generalizability Theory*. New York, NY: Springer-Verlag.
- Crocker, L., and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Belmont, CA: Wadsworth Group.
- Kolen, M. J., and Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2<sup>nd</sup> Ed.). New York, NY: Springer Science and Business Media, LLC.
- Koretz, D. M., and Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational Measurement* (4<sup>th</sup> ed.) (pp. 531–578). Westport, CT: American Council on Education and Praeger Publishers.
- Kuder, G. F., and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika* 2:151–160.
- Livingston, S. A., and Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement* 32(2):179–197.
- Qualls, L. A. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education* 8(2):111–120.

## **Appendices**

---

Appendix A: Sample OAE Candidate Score Reports

Appendix B: Test Form Statistics Report for Test Fields with More than 100 Tests Taken

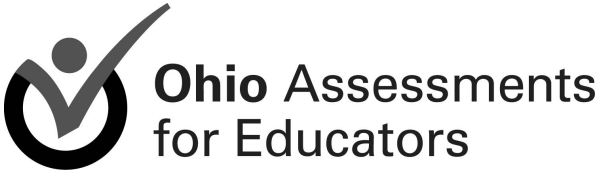
Appendix C: Test Form Statistics Report for Test Fields with Fewer than 100 Tests Taken

Appendix D: Test Field Statistics Report for Test Fields with Constructed-Response Items

Appendix E: Pass Rate Summary

## **Appendix A: Sample OAE Candidate Score Reports: Pass and Did Not Pass**

Note: Candidate Score Reports include only the last five digits of candidate social security numbers with four leading X's, (e.g., XXX-X1-2345)



FirstName LastName  
300 Venture Way  
Sample City, OH XXXXX

<b>Assessment of Professional Knowledge: Adolescence to Young Adult (7-12) 003</b> <b>Minimum Passing Score: 220</b> <b>Assessment Date: August 20, 2024</b>	
<b>Status: Pass</b>	<b>Your Score: 231</b>

Domain	Performance Index
Student Development and Learning	++
Assessment, Instruction, and the Learning Environment	+++
The Professional Environment	+++
Constructed Response Assignments	+++

Examinee Name: FirstName LastName

Social Security Number: XXX-XX-XXXX



This barcode contains unique candidate information.



## HOW TO READ YOUR SCORE REPORT

**Overview.** This score report provides your assessment results for the Ohio Assessments for Educators: Initial Licensure (OAE: Initial Licensure) that you took on the assessment date indicated on the report. The purpose of the OAE: Initial Licensure is to assess the knowledge and skills of a prospective Ohio educator. The OAE: Initial Licensure program helps the Ohio Department of Education and the State Board of Education of Ohio meet their goal of ensuring that candidates have the knowledge and skills needed to perform the job of an educator in Ohio schools.

**Assessment Score.** Your total assessment score is a scaled score. A scaled score is a combination of the number of scorable questions you answered correctly on the multiple-choice section of the assessment and the scores you received on any constructed-response assignments (if applicable) converted to a scale from 100 to 300, with a score of 220 representing the passing score for the assessment. The number of questions per competency may vary by test form. The total number of questions needed to pass the assessment may also vary across test forms as a result of variations in question difficulty. The conversion of raw scores to scaled scores takes such differences into account. The passing score for each assessment is established by the State Board of Education of Ohio and is based on the professional judgments and recommendations of Ohio educators. "Pass" or "Did Not Pass" status is based on your total score for each assessment.

**Performance Indices.** Performance indices are provided for each domain of the assessment. These indices can help you understand your areas of strength and weakness. This information should be interpreted with caution, since different domains contain different numbers of assessment questions. For each domain, your performance is reported according to the following performance indices:

Performance Indices: Multiple-Choice Questions	
++++	Performance on the competencies included in the domain is well above the level represented by the minimum passing score.
+++	Performance on the competencies included in the domain is just at or above the level represented by the minimum passing score.
++	Performance on the competencies included in the domain is just below the level represented by the minimum passing score.
+	Performance on the competencies included in the domain is well below the level represented by the minimum passing score.

Performance Indices: Constructed-Response Assignments			
Constructed-Response Assignments (Assessments with a 4-Point Scoring Scale*)		Constructed-Response Assignments (Assessments with a 3-Point Scoring Scale*)	
++++	Response reflects a thorough understanding	+++	Response reflects a thorough understanding
+++	Response reflects a general understanding	++	Response reflects a general understanding
++	Response reflects a limited understanding	+	Response reflects limited or no understanding
+	Response reflects little or no understanding		

\*See study guides at [www.oh.educatorstests.com](http://www.oh.educatorstests.com)





Each assessment framework describes the content knowledge assessed by the OAE: Initial Licensure. You may view, print, or download the framework for any OAE: Initial Licensure assessment by selecting "Assessment Frameworks" in the "Prepare" section of the OAE: Initial Licensure website at [www.oh.educatortests.com](http://www.oh.educatortests.com).

If a response to a constructed-response assignment is designated "Blank" or "Unscorable," you will see one of the following codes:

Codes for Blank/Unscorable Responses	
<b>B</b>	Response was blank
<b>U-1</b>	Response was unrelated to assigned topic
<b>U-2</b>	Response was unintelligible
<b>U-3</b>	Response was not primarily in English
<b>U-4</b>	Response lacked sufficient amount of original work

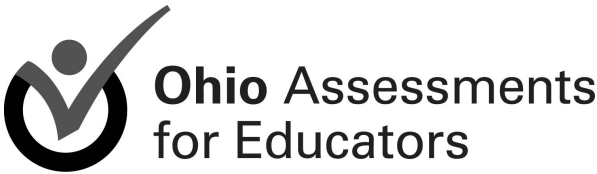
**Summary Report for Assessments with Two Subtests.** Some OAE: Initial Licensure assessments are comprised of two subtests. If you took a two-subtest assessment, your score report will include an assessment summary. The assessment summary lists passing status information for each subtest within the assessment and includes the date on which you passed each subtest. Please note that for assessments composed of more than one subtest, you must pass both subtests for that assessment to meet Ohio licensure requirements.

**Reporting of Scores.** Your scores are reported directly to the Ohio Department of Education and the Ohio educator preparation institution(s) you indicated during the registration process. This score report is for your information only. Keep a copy for your permanent records.

#### ADDITIONAL INFORMATION

**Retaking the Assessment.** You may retake an OAE: Initial Licensure assessment by following the same registration procedures you completed for previous assessment administrations. For information about retake policies, visit [www.oh.educatortests.com](http://www.oh.educatortests.com).

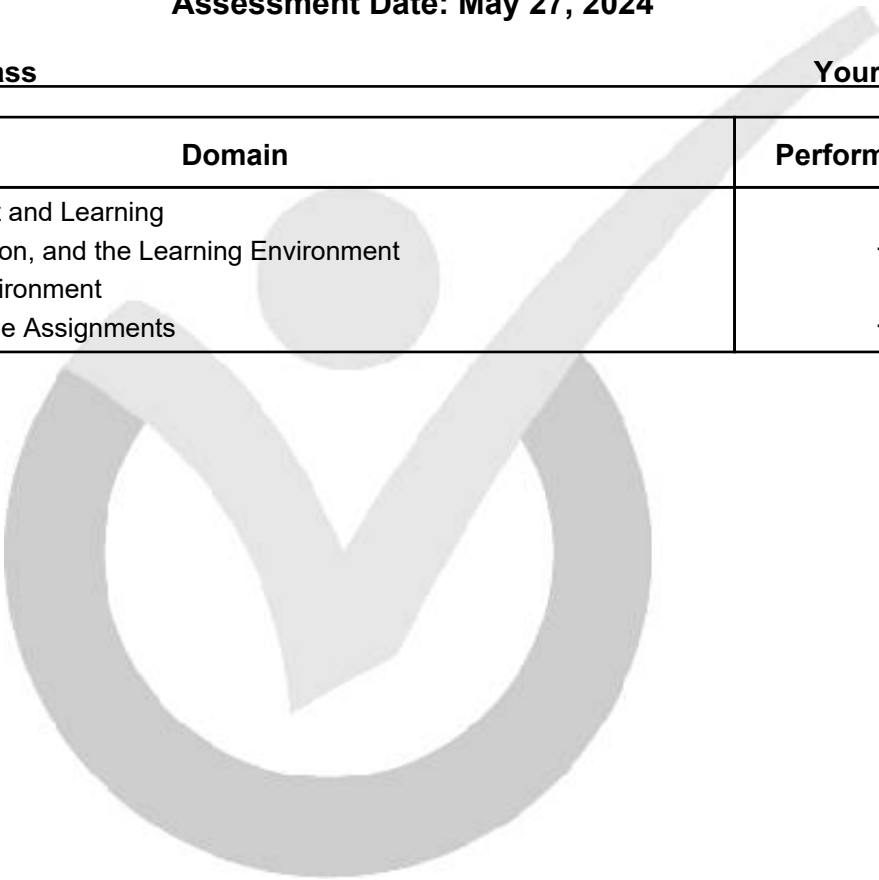
**Assessment Preparation Materials.** Assessment preparation materials are available through the website you used to register for this assessment.



FirstName LastName  
300 Venture Way  
Sample City, OH XXXXX

<b>Assessment of Professional Knowledge: Multi-Age (PK-12) (004)</b> <b>Minimum Passing Score: 220</b> <b>Assessment Date: May 27, 2024</b>	
<b>Status: Did Not Pass</b>	<b>Your Score: 207</b>

Domain	Performance Index
Student Development and Learning	+
Assessment, Instruction, and the Learning Environment	+++
The Professional Environment	+
Constructed Response Assignments	+++



Examinee Name: FirstName LastName

Social Security Number: XXX-XX-XXXX



This barcode contains unique candidate information.



## HOW TO READ YOUR SCORE REPORT

**Overview.** This score report provides your assessment results for the Ohio Assessments for Educators: Initial Licensure (OAE: Initial Licensure) that you took on the assessment date indicated on the report. The purpose of the OAE: Initial Licensure is to assess the knowledge and skills of a prospective Ohio educator. The OAE: Initial Licensure program helps the Ohio Department of Education and the State Board of Education of Ohio meet their goal of ensuring that candidates have the knowledge and skills needed to perform the job of an educator in Ohio schools.

**Assessment Score.** Your total assessment score is a scaled score. A scaled score is a combination of the number of scorable questions you answered correctly on the multiple-choice section of the assessment and the scores you received on any constructed-response assignments (if applicable) converted to a scale from 100 to 300, with a score of 220 representing the passing score for the assessment. The number of questions per competency may vary by test form. The total number of questions needed to pass the assessment may also vary across test forms as a result of variations in question difficulty. The conversion of raw scores to scaled scores takes such differences into account. The passing score for each assessment is established by the State Board of Education of Ohio and is based on the professional judgments and recommendations of Ohio educators. "Pass" or "Did Not Pass" status is based on your total score for each assessment.

**Performance Indices.** Performance indices are provided for each domain of the assessment. These indices can help you understand your areas of strength and weakness. This information should be interpreted with caution, since different domains contain different numbers of assessment questions. For each domain, your performance is reported according to the following performance indices:

Performance Indices: Multiple-Choice Questions	
++++	Performance on the competencies included in the domain is well above the level represented by the minimum passing score.
+++	Performance on the competencies included in the domain is just at or above the level represented by the minimum passing score.
++	Performance on the competencies included in the domain is just below the level represented by the minimum passing score.
+	Performance on the competencies included in the domain is well below the level represented by the minimum passing score.

Performance Indices: Constructed-Response Assignments			
Constructed-Response Assignments (Assessments with a 4-Point Scoring Scale*)		Constructed-Response Assignments (Assessments with a 3-Point Scoring Scale*)	
++++	Response reflects a thorough understanding	+++	Response reflects a thorough understanding
+++	Response reflects a general understanding	++	Response reflects a general understanding
++	Response reflects a limited understanding	+	Response reflects limited or no understanding
+	Response reflects little or no understanding		

\*See study guides at [www.oh.educatorstests.com](http://www.oh.educatorstests.com)



Each assessment framework describes the content knowledge assessed by the OAE: Initial Licensure. You may view, print, or download the framework for any OAE: Initial Licensure assessment by selecting "Assessment Frameworks" in the "Prepare" section of the OAE: Initial Licensure website at [www.oh.educatorstests.com](http://www.oh.educatorstests.com).

If a response to a constructed-response assignment is designated "Blank" or "Unscorable," you will see one of the following codes:

Codes for Blank/Unscorable Responses	
<b>B</b>	Response was blank
<b>U-1</b>	Response was unrelated to assigned topic
<b>U-2</b>	Response was unintelligible
<b>U-3</b>	Response was not primarily in English
<b>U-4</b>	Response lacked sufficient amount of original work

**Summary Report for Assessments with Two Subtests.** Some OAE: Initial Licensure assessments are comprised of two subtests. If you took a two-subtest assessment, your score report will include an assessment summary. The assessment summary lists passing status information for each subtest within the assessment and includes the date on which you passed each subtest. Please note that for assessments composed of more than one subtest, you must pass both subtests for that assessment to meet Ohio licensure requirements.

**Reporting of Scores.** Your scores are reported directly to the Ohio Department of Education and the Ohio educator preparation institution(s) you indicated during the registration process. This score report is for your information only. Keep a copy for your permanent records.

#### ADDITIONAL INFORMATION

**Retaking the Assessment.** You may retake an OAE: Initial Licensure assessment by following the same registration procedures you completed for previous assessment administrations. For information about retake policies, visit [www.oh.educatorstests.com](http://www.oh.educatorstests.com).

**Assessment Preparation Materials.** Assessment preparation materials are available through the website you used to register for this assessment.

## Detailed Performance Summary

**Examinee Name:** FirstName LastName

**Assessment Date:** May 29, 2023

**Assessment Name/Code:** Assessment of Professional Knowledge: Multi-Age (PK-12) (004)

### Multiple-Choice Performance

The multiple-choice section of the assessment contributed 80% of your total assessment score.

This section provides information about the number and percentage of questions answered correctly in each domain and competency.

Domain and Competency	No. of Scorable Questions on Assessment	No. of Questions Answered Correctly	Percent Correct
<b>Student Development and Learning</b>	<b>24</b>	<b>13</b>	<b>54%</b>
0001 Processes of human development	8	3	38%
0002 Learning processes	8	5	63%
0003 Student diversity	8	5	63%
<b>Assessment, Instruction, and the Learning Environment</b>	<b>40</b>	<b>29</b>	<b>73%</b>
0004 Assessment instruments and practices	8	7	88%
0005 Curricular and instructional planning	8	8	100%
0006 Various instructional approaches	8	4	50%
0007 Motivation and communication	8	6	75%
0008 Structure and manage the classroom	8	4	50%
<b>The Professional Environment</b>	<b>16</b>	<b>8</b>	<b>50%</b>
0009 Establish partnerships and collaborate	8	6	75%
0010 Roles and expectations for professional educators	8	2	25%

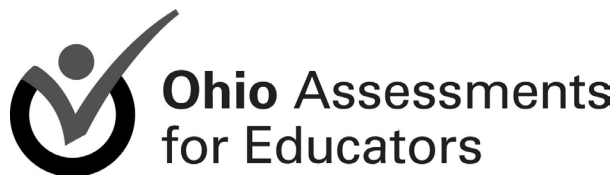
Note: Results for competencies with fewer than five scorable questions should be read with caution. The number of questions per competency will vary by test form. The total number of questions needed to pass the test may also vary across test forms as a result of variations in question difficulty. The conversion of raw scores to scaled scores takes such differences into account.

Examinee Name: FirstName LastName

Social Security Number: XXX-XX-XXXX



This barcode contains unique candidate information.



## Detailed Performance Summary

**Examinee Name:** FirstName LastName

**Assessment Date:** May 27, 2024

**Assessment Name/Code:** Assessment of Professional Knowledge: Multi-Age (PK-12) (004)

### Constructed-Response Performance

The constructed-response section of the assessment contributed 20% of your total assessment score.

This section provides information about your performance for each constructed-response assignment.

#### Written Assignment: Case Study:

Maximum number of points for the assignment: 8

Total number of points you received for the assignment: 7

Your response reflects thorough or general understanding of the relevant knowledge and skills as defined in the relevant OAE framework objectives. The response reflects one or more of the following characteristics.

- The response thoroughly or largely achieves the purpose of the assignment.
- The response demonstrates an accurate and effective or generally accurate and effective application of relevant content knowledge.
- The response provides relevant support.
- The response reflects sound or generally sound reasoning about the topic.

#### Written Assignment: Work Product:

Maximum number of points for the assignment: 8

Total number of points you received for the assignment: 6

Your response reflects a general understanding of the relevant knowledge and skills as defined in the relevant OAE framework objectives.

- The response largely achieves the purpose of the assignment.
- The response demonstrates a generally accurate and generally effective application of relevant content knowledge.
- The response provides some relevant support.
- The response reflects generally sound reasoning about the topic.

Examinee Name: FirstName LastName

Social Security Number: XXX-XX-XXXX



This barcode contains unique candidate information.

## **Appendix B: Test Form Statistics Report for Test Fields with More than 100 Tests Taken**

*Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form*

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
001 Assessment of Professional Knowledge: Early Childhood (PK-3)	A	27	226	--	--	--	80	--	2	--
	B	25	230	--	--	--	80	--	2	--
	C	24	228	--	--	--	80	--	2	--
	D	18	226	--	--	--	80	--	2	--
	E	15	219	--	--	--	80	--	2	--
	F	13	219	--	--	--	80	--	2	--
	G	12	225	--	--	--	80	--	2	--
	H	12	229	--	--	--	80	--	2	--
	J	8	--	--	--	--	80	--	2	--
	K	7	--	--	--	--	80	--	2	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)



*Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form*

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
002 Assessment of Professional Knowledge: Middle Childhood (4-9)	A	80	236	--	--	--	80	--	2	--
	B	79	236	--	--	--	80	--	2	--
	C	70	242	--	--	--	80	--	2	--
	D	69	243	--	--	--	80	--	2	--
	E	15	232	--	--	--	80	--	2	--
	F	13	245	--	--	--	80	--	2	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

*Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form*

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
003 Assessment of Professional Knowledge: Adolescence to Young Adult (7-12)	A	151	244	11.6	0.94	0.81	80	0.80	2	0.30
	B	137	250	11.8	0.95	0.86	80	0.85	2	0.56
	C	60	248	--	--	--	80	--	2	--
	D	57	250	--	--	--	80	--	2	--
	E	52	250	--	--	--	80	--	2	--
	F	49	253	--	--	--	80	--	2	--
	G	44	247	--	--	--	80	--	2	--
	H	39	247	--	--	--	80	--	2	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
004 Assessment of Professional Knowledge: Multi-Age (PK-12)	A	163	248	13.1	0.93	0.79	80	0.78	2	0.37
	B	157	249	12.9	0.94	0.87	80	0.87	2	0.46
	C	145	244	12.1	0.91	0.88	80	0.88	2	0.42
	D	142	244	12.9	0.91	0.84	80	0.84	2	0.42
	E	135	248	12.2	0.94	0.86	80	0.85	2	0.48
	F	128	249	12.7	0.93	0.84	80	0.84	2	0.38
	G	120	246	12.5	0.94	0.89	80	0.89	2	0.37
	H	118	250	13.2	0.92	0.85	80	0.84	2	0.42
	J	91	239	--	--	--	80	--	2	--
	K	81	246	--	--	--	80	--	2	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
006 Art	A	166	230	9.7	0.86	--	120	0.87	--	--
	B	148	227	10.2	0.79	--	120	0.89	--	--
007 Biology	A	98	208	--	--	--	120	--	--	--
	B	76	212	--	--	--	120	--	--	--
008 Business Education	A	53	225	--	--	--	120	--	--	--
	B	51	224	--	--	--	120	--	--	--
012 Early Childhood Education	A	31	233	--	--	--	120	--	--	--
	B	28	227	--	--	--	120	--	--	--
	C	18	219	--	--	--	120	--	--	--
	D	14	229	--	--	--	120	--	--	--
	E	14	218	--	--	--	120	--	--	--
	F	14	234	--	--	--	120	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
012 Early Childhood Education	G	13	221	--	--	--	120	--	--	--
	H	7	--	--	--	--	120	--	--	--
013 Early Childhood Special Education	A	65	237	--	--	--	80	--	--	--
	B	59	233	--	--	--	80	--	--	--
	C	49	231	--	--	--	80	--	--	--
	D	46	229	--	--	--	80	--	--	--
	E	13	247	--	--	--	80	--	--	--
	F	13	216	--	--	--	80	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
015 Educational Leadership	A	236	230	10.4	0.87	0.74	60	0.73	4	0.51
	B	207	233	10.6	0.91	0.79	60	0.75	4	0.62
	C	206	230	10.4	0.84	0.74	60	0.74	4	0.51
	D	196	231	10.4	0.87	0.75	60	0.77	4	0.45
	E	64	233	--	--	--	60	--	4	--
	F	64	236	--	--	--	60	--	4	--
018 Elementary Education Subtest I	A	205	223	12.5	0.80	--	60	0.80	--	--
	B	185	223	12.1	0.80	--	60	0.80	--	--
	C	13	225	--	--	--	60	--	--	--
	D	13	207	--	--	--	60	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

*Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form*

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consist-ency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
019 Elementary Education Subtest II	A	111	208	14.3	0.84	--	60	0.79	--	--
	B	108	212	14.0	0.76	--	60	0.70	--	--
	C	100	215	13.4	0.77	--	60	0.78	--	--
	D	93	210	--	--	--	60	--	--	--
	E	83	216	--	--	--	60	--	--	--
	F	80	212	--	--	--	60	--	--	--
	G	19	199	--	--	--	60	--	--	--
	H	17	193	--	--	--	60	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
020 English Language Arts	A	276	228	9.3	0.89	--	120	0.92	--	--
	B	247	227	8.8	0.89	--	120	0.92	--	--
	C	55	228	--	--	--	120	--	--	--
	D	40	226	--	--	--	120	--	--	--
021 English to Speakers of Other Languages (ESOL)	A	173	244	8.1	0.95	--	120	0.92	--	--
	B	158	248	7.9	0.95	--	120	0.92	--	--
	C	22	242	--	--	--	120	--	--	--
	D	14	246	--	--	--	120	--	--	--
023 Health	A	84	242	--	--	--	80	--	--	--
	B	77	237	--	--	--	80	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)



**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
024 Integrated Science	A	96	236	--	--	--	120	--	--	--
	B	86	238	--	--	--	120	--	--	--
	C	19	230	--	--	--	120	--	--	--
	D	14	231	--	--	--	120	--	--	--
025 Integrated Social Studies	A	357	227	7.3	0.85	--	120	0.92	--	--
	B	353	223	7.5	0.87	--	120	0.91	--	--
	C	68	224	--	--	--	120	--	--	--
	D	61	226	--	--	--	120	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
027 Mathematics	A	101	217	7.7	0.85	--	120	0.96	--	--
	B	95	220	--	--	--	120	--	--	--
	C	85	216	--	--	--	120	--	--	--
	D	81	220	--	--	--	120	--	--	--
	E	76	218	--	--	--	120	--	--	--
	F	54	219	--	--	--	120	--	--	--
	G	22	213	--	--	--	120	--	--	--
	H	18	203	--	--	--	120	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consist-ency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
028 Middle Grades English Language Arts	A	159	241	9.0	0.88	--	100	0.88	--	--
	B	153	239	9.5	0.91	--	100	0.87	--	--
	C	19	239	--	--	--	100	--	--	--
	D	19	233	--	--	--	100	--	--	--
029 Middle Grades Science	A	157	231	7.1	0.83	--	100	0.86	--	--
	B	154	235	7.2	0.87	--	100	0.89	--	--
	C	21	225	--	--	--	100	--	--	--
	D	14	252	--	--	--	100	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
030 Middle Grades Mathematics	A	211	233	7.8	0.88	--	100	0.90	--	--
	B	189	228	8.2	0.86	--	100	0.91	--	--
	C	19	223	--	--	--	100	--	--	--
	D	13	224	--	--	--	100	--	--	--
031 Middle Grades Social Studies	A	274	224	7.8	0.75	--	100	0.87	--	--
	B	253	221	8.2	0.83	--	100	0.84	--	--
	C	28	225	--	--	--	100	--	--	--
	D	27	225	--	--	--	100	--	--	--
032 Music	A	242	228	10.1	0.78	--	120	0.87	--	--
	B	231	222	10.6	0.82	--	120	0.88	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
034 Physical Education	A	166	226	9.4	0.79	--	120	0.86	--	--
	B	161	226	9.7	0.84	--	120	0.85	--	--
	C	27	208	--	--	--	120	--	--	--
	D	24	218	--	--	--	120	--	--	--
038 Reading Subtest I	A	80	263	--	--	--	54	--	2	--
	B	77	260	--	--	--	54	--	2	--
	C	72	255	--	--	--	54	--	2	--
	D	67	254	--	--	--	54	--	2	--
	E	6	--	--	--	--	54	--	2	--
	F	4	--	--	--	--	54	--	2	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
039 Reading Subtest II	A	90	262	--	--	--	54	--	2	--
	B	76	250	--	--	--	54	--	2	--
	C	68	254	--	--	--	54	--	2	--
	D	60	261	--	--	--	54	--	2	--
	E	6	--	--	--	--	54	--	2	--
	F	5	--	--	--	--	54	--	2	--
040 School Counselor	A	143	221	10.9	0.84	--	120	0.88	--	--
	B	140	224	11.0	0.79	--	120	0.85	--	--
	C	23	208	--	--	--	120	--	--	--
	D	22	220	--	--	--	120	--	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

**Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form**

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
043 Special Education	A	315	225	10.3	0.83	--	120	0.88	--	--
	B	297	227	10.3	0.84	--	120	0.87	--	--
	C	293	229	9.7	0.88	--	120	0.90	--	--
	D	291	228	9.6	0.85	--	120	0.89	--	--
	E	269	232	9.3	0.88	--	120	0.91	--	--
	F	259	226	10.3	0.84	--	120	0.89	--	--
	G	137	223	9.1	0.91	--	120	0.94	--	--
	H	131	225	9.6	0.85	--	120	0.90	--	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)

***Test Form Statistics Report For Test Fields With More Than 100 Tests Taken Per Test Form***

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Indices				Multiple-Choice Section		Open Response Section	
			Mean <sup>a</sup>	SEM <sup>b</sup>	Decision Consistency <sup>c</sup>	Stratified Alpha <sup>d</sup>	Length <sup>e</sup>	KR20 <sup>f</sup>	Length <sup>e</sup>	G Coefficient <sup>g</sup>
Test Field *	Form **									
053 Gifted Education	A	78	234	--	--	--	100	--	--	--
	B	76	245	--	--	--	100	--	--	--
055 Primary Education (PK-5)	A	2,655	241	8.4	0.90	--	100	0.85	--	--
057 Assessment of Professional Knowledge: Primary Education (PK-5)	A	1,250	257	10.2	0.99	0.83	80	0.82	2	0.39
058 Primary Special Education (PK-5)	A	583	244	8.3	0.94	--	100	0.81	--	--
190 Foundations of Reading	A	6,402	225	10.8	0.84	0.88	85	0.88	2	0.40

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively

*a* - the mean total test scaled scores achieved by the candidates taking the test form

*b* - the standard error of measurement of the total test scaled scores

*c* - a reliability statistic that describes the consistency of the pass/fail decision on the total test scaled score

*d* - an estimate of total test reliability for a test containing multiple item types (e.g., multiple-choice, constructed-response items)

*e* - the number of scorable items in the multiple-choice or open-response section

*f* - an overall test consistency (reliability) estimate based on a single test administration (Kuder-Richardson formula 20)

*g* - a measure of the proportion of total score variance that is attributable to true score variability rather than to measurement error (Generalizability Coefficient)



## **Appendix C: Test Form Statistics Report for Test Fields with Fewer than 100 Tests Taken**

***Test Form Statistics Report For Test Fields With Fewer Than 100 Tests Taken Per Test Form***

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Mean <sup>a</sup>	Multiple-Choice Section Length <sup>b</sup>	Open-Response Section Length <sup>b</sup>
Test Field *	Form **				
005 Agriscience	A	20	230	80	--
	B	20	233	80	--
009 Chemistry	A	18	237	120	--
	B	18	227	120	--
011 Dance	A	9	--	80	--
014 Earth and Space Science	A	17	221	120	--
	B	15	220	120	--
016 Computer/Technology Subtest I	A	25	255	56	2
	B	15	248	56	2
017 Computer/Technology Subtest II	A	18	268	32	3
	B	16	258	32	3
022 Family and Consumer Sciences	A	38	228	120	--
	B	30	219	120	--
026 Marketing	A	4	--	80	--
035 Physics	A	18	247	120	--
	B	13	226	120	--

(Continued)

***NOTES: \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively  
a - the mean total test scaled scores achieved by the candidates taking the test form  
b - the number of scorable items in the multiple-choice or open-response section***

***Test Form Statistics Report For Test Fields With Fewer Than 100 Tests Taken Per Test Form***

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Mean <sup>a</sup>	Multiple-Choice Section Length <sup>b</sup>	Open-Response Section Length <sup>b</sup>
Test Field *	Form **				
036 Prekindergarten Subtest I	A	28	233	54	2
	B	23	226	54	2
037 Prekindergarten Subtest II	A	25	233	54	2
	B	23	236	54	2
041 School Library Media Specialist	A	35	215	120	--
	B	27	219	120	--
044 Special Education Specialist: Deaf/Hard of Hearing	A	18	242	80	--
045 Special Education Specialist: Visually Impaired	A	14	223	64	1
	B	11	223	64	1
046 Technology Education Subtest I	A	11	217	36	3
047 Technology Education Subtest II	A	15	209	60	3
048 Theater	A	34	225	85	2
050 American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest I)	A	24	221	40	--
051 American Sign Language (ASL) Assessment for World Language Teachers of ASL (Subtest II)	A	20	218	40	4
054 Computer Science	A	60	240	100	--
059 Financial Literacy	A	42	239	32	--

(Continued)

**NOTES:** \*, \*\*, \*\*\* - test field, test form administered, and the number of times the test form was taken, respectively  
<sup>a</sup> - the mean total test scaled scores achieved by the candidates taking the test form  
<sup>b</sup> - the number of scorable items in the multiple-choice or open-response section

***Test Form Statistics Report For Test Fields With Fewer Than 100 Tests Taken Per Test Form***

September 1, 2023 - August 31, 2024		Number of Tests Taken ***	Total Test Scaled Score Mean <sup>a</sup>	Multiple- Choice Section Length <sup>b</sup>	Open- Response Section Length <sup>b</sup>
Test Field *	Form **				
060 Dual Middle Childhood Special Education (4-9)	A	7	--	60	--
061 Dual Adolescence to Young Adult Special Education (7-12)	A	6	--	60	--

***NOTES: \****, \*\*, \*\*\* - *test field, test form administered, and the number of times the test form was taken, respectively*  
*a* - *the mean total test scaled scores achieved by the candidates taking the test form*  
*b* - *the number of scorable items in the multiple-choice or open-response section*

## **Appendix D: Test Field Statistics Report for Test Fields with Constructed-Response Items**

Sep 1, 2023 - Aug 31, 2024 Test and Items		Number of Tests Taken <sup>a</sup>	Number of Valid Scores <sup>b</sup>	Open Response Prompts			
				Scorer Agreement (%)			Inter- rater Reliabil- ity <sup>f</sup>
				Agree- ment <sup>c</sup>	Exact <sup>d</sup>	Adja- cent <sup>e</sup>	
001 Assessment of Professional Knowledge: Early Childhood (PK-3)	1	161	161	99.4	93.8	5.6	0.94
	2	161	158	100.0	93.0	7.0	0.96
002 Assessment of Professional Knowledge: Middle Childhood (4-9)	1	326	325	99.7	82.8	16.9	0.85
	2	326	325	100.0	80.0	20.0	0.90
003 Assessment of Professional Knowledge: Adolescence to Young Adult (7-12)	1	589	587	99.8	73.6	26.2	0.76
	2	589	587	99.8	75.8	24.0	0.79
004 Assessment of Professional Knowledge: Multi-Age (PK-12)	1	1,280	1,270	99.8	80.0	19.8	0.82
	2	1,280	1,261	99.7	77.8	21.9	0.85
015 Educational Leadership	1	973	971	99.2	71.1	28.1	0.65
	2	973	970	99.3	73.1	26.2	0.71
	3	973	968	98.6	69.4	29.1	0.78
	4	973	962	100.0	80.1	19.9	0.76
038 Reading Subtest I	1	306	305	98.7	83.3	15.4	0.90
	2	306	305	100.0	78.0	22.0	0.88
039 Reading Subtest II	1	305	305	100.0	86.9	13.1	0.91
	2	305	304	100.0	86.5	13.5	0.92
057 Assessment of Professional Knowledge: Primary Education (PK-5)	1	1,250	1,236	99.4	71.3	28.1	0.69
	2	1,250	1,232	99.2	74.4	24.8	0.63

(Continued)

**NOTES:** *a - the number of times the test was taken*  
*b - the number of scorable responses included in the calculations*  
*c - overall agreement determined by summing exact and adjacent agreement*  
*d - the percentage of scores in which the first two scorers were in exact agreement*  
*e - the percentage of scores in which the two scorers assigned adjacent scores*  
*f - the intraclass correlation between the first and second score assigned to each response, corrected using the Spearman-Brown formula*

Sep 1, 2023 - Aug 31, 2024 Test and Items		Number of Tests Taken <sup>a</sup>	Number of Valid Scores <sup>b</sup>	Open Response Prompts			
				Scorer Agreement (%)			Inter- rater Reliabil- ity <sup>f</sup>
				Agree- ment <sup>c</sup>	Exact <sup>d</sup>	Adja- cent <sup>e</sup>	
190 Foundations of Reading	1	6,402	6,365	99.8	79.0	20.7	0.83
	2	6,402	6,345	99.6	72.7	26.9	0.78

**NOTES:** *a - the number of times the test was taken*  
*b - the number of scorable responses included in the calculations*  
*c - overall agreement determined by summing exact and adjacent agreement*  
*d - the percentage of scores in which the first two scorers were in exact agreement*  
*e - the percentage of scores in which the two scorers assigned adjacent scores*  
*f - the intraclass correlation between the first and second score assigned to each response, corrected using the Spearman-Brown formula*

## **Appendix E: Pass Rate Summary**



Ohio Assessments for Educators  
PASS RATE ANALYSIS - September 1, 2023-August 31, 2024

OAE Test Field	Number of Examinees & Their Percent at or Above State Passing Score							
	Initial Attempt		Initial Attempt		Initial Attempt		Best Attempt	
	9/1/21 to 8/31/22		9/1/22 to 8/31/23		9/1/23 to 8/31/24		9/1/21 to 8/31/24	
	N	%	N	%	N	%	N	%
001 Assessment of Professional Knowledge: Early Childhood (PK-3)	1089	86%	928	86%	99	79%	2191	91%
057 Assessment of Professional Knowledge: Primary Education (PK-5)	0	0%	105	99%	1228	98%	1333	99%
002 Assessment of Professional Knowledge: Middle Childhood (4-9)	300	86%	231	84%	270	82%	833	91%
003 Assessment of Professional Knowledge: Adoles. to Young Adult (7-12)	649	96%	598	94%	548	94%	1814	97%
004 Assessment of Professional Knowledge: Multi-Age (PK-12)	1105	90%	1086	91%	1103	91%	3396	95%
005 Agriscience	28	89%	22	100%	28	71%	79	95%
006 Art	181	77%	160	78%	232	77%	597	89%
007 Biology	98	48%	104	50%	87	43%	321	72%
008 Business Education	48	77%	53	68%	74	65%	181	77%
009 Chemistry	32	69%	35	63%	20	75%	90	82%
011 Dance	9	89%	12	67%	8	88%	29	86%
012 Early Childhood Education	969	83%	562	82%	79	81%	1706	91%
013 Early Childhood Special Education	610	73%	538	76%	176	82%	1375	86%
014 Earth and Space Science	16	38%	16	50%	22	50%	58	74%
015 Educational Leadership	898	86%	834	85%	787	82%	2631	90%
016 Computer/Technology Subtest I	81	95%	44	95%	35	80%	160	98%
017 Computer/Technology Subtest II	80	95%	45	98%	34	97%	161	98%
018 Elementary Education Subtest I	1683	68%	1930	68%	192	72%	3944	84%
019 Elementary Education Subtest II	1660	57%	1928	56%	195	64%	4028	77%
020 English Language Arts	507	80%	395	76%	439	75%	1395	85%
021 English to Speakers of Other Languages (ESOL)	317	91%	293	90%	330	89%	961	92%
022 Family and Consumer Sciences	41	83%	47	79%	52	67%	145	83%
023 Health	126	77%	133	87%	136	88%	407	90%
024 Integrated Science	191	80%	166	80%	162	79%	536	88%
025 Integrated Social Studies	448	70%	434	69%	539	69%	1504	82%
026 Marketing	2	0%	3	100%	4	75%	9	67%
027 Mathematics	294	55%	271	58%	303	55%	926	73%
028 Middle Grades Eng. Language Arts	343	85%	319	88%	308	84%	991	92%
029 Middle Grades Science	301	75%	246	81%	267	78%	858	91%
030 Middle Grades Mathematics	410	72%	275	72%	308	70%	1049	85%
031 Middle Grades Social Studies	347	65%	287	71%	345	61%	1055	82%
032 Music	314	77%	270	70%	284	76%	899	89%
034 Physical Education	181	70%	205	68%	260	69%	676	81%
035 Physics	21	76%	18	83%	25	80%	69	86%
036 Prekindergarten Subtest I	59	92%	38	84%	39	77%	143	84%
037 Prekindergarten Subtest II	58	83%	39	82%	37	84%	138	87%
038 Reading Subtest I	410	96%	334	95%	294	96%	1052	97%
039 Reading Subtest II	411	94%	325	95%	285	96%	1033	97%

Ohio Assessments for Educators  
PASS RATE ANALYSIS - September 1, 2023-August 31, 2024

62

OAE Test Field	Number of Examinees & Their Percent at or Above State Passing Score							
	Initial Attempt		Initial Attempt		Initial Attempt		Best Attempt	
	9/1/21 to 8/31/22		9/1/22 to 8/31/23		9/1/23 to 8/31/24		9/1/21 to 8/31/24	
	N	%	N	%	N	%	N	%
040 School Counselor	313	76%	285	73%	196	73%	856	86%
041 School Library Media Specialist	35	63%	31	68%	39	64%	111	76%
042 School Psychologist	2	100%	1	100%	0	0%	3	100%
043 Special Education	1388	77%	1308	76%	1390	75%	4263	86%
044 Special Education Specialist: Deaf/Hard of Hearing	17	88%	20	75%	18	89%	55	89%
045 Special Education Specialist: Visually Impaired	11	82%	12	58%	19	53%	48	75%
046 Technology Education Subtest I	11	55%	13	85%	9	56%	33	70%
047 Technology Education Subtest II	11	36%	12	50%	9	22%	34	59%
048 Theater	19	84%	31	81%	25	64%	77	81%
050 American Sign Language (Subtest I)	21	62%	24	79%	15	67%	61	89%
051 American Sign Language (Subtest II)	18	100%	16	100%	14	57%	48	88%
053 Gifted Education	135	84%	157	89%	137	85%	441	93%
054 Computer Science	34	91%	50	62%	51	80%	139	83%
055 Primary Education (PK-5)	0	0%	222	86%	2392	89%	2614	94%
058 Primary Special Education (PK-5)	0	0%	63	92%	556	94%	619	96%
059 Financial Literacy	0	0%	0	0%	41	80%	41	80%
060 Dual Middle Childhood Special Education (4-9)	0	0%	0	0%	7	100%	7	100%
061 Dual Adolescence to Young Adult Special Education (7-12)	0	0%	0	0%	6	33%	6	33%
090 Foundations of Reading (retired 1/2023)	4159	75%	1365	73%	0	0%	6101	84%
190 Foundations of Reading (first test date 1/2023)	0	0%	3213	65%	4271	69%	7484	82%

**Developed and Produced by**



**Evaluation Systems  
300 Venture Way, Hadley, MA 01035**