

ATTACHMENT A

This Solicitation is a Contract Document and is a request for proposal in connection with the Contract awarded by the Office of Management and Enterprise Services, as more particularly described below. Any defined term used herein but not defined herein shall have the meaning ascribed in the General Terms or other Contract Document.

Purpose

The Contract is awarded on behalf of the Oklahoma Office of Educational Quality and Accountability (OEQA) for the provision of developing and implementing a comprehensive assessment of general and subject matter knowledge and skills of educator candidates.

1. Contract Term and Renewal Options

The Contract term, which begins on September 1, 2021, shall be effective for one (1) year with an option to renew annually for five (5) additional years with the same terms and conditions until August 31, 2027.

2. Ongoing Program Reporting Requirements and Tentative Schedule.

- 2.1 The initial product from the supplier shall be the development of a detailed plan for accomplishing each requirement of the testing program outlined in this RFP. The implementation schedule for project shall follow the outline as listed, but the supplier and OEQA shall work together to set dates so that the first test administration under the contract shall be in **September 2021**.
- 2.2 The supplier shall provide monthly reports by the first business day of the month to OEQA of progress toward the implementation of the general education test and subject matter tests. These reports shall indicate: meetings held, upcoming meetings, areas of problems that may cause delays in schedules, areas in which activities are ahead of or behind schedule, and any other relevant information that shall assist the OEQA in monitoring the activities of the supplier.
- 2.3 The implementation schedule is proposed as follows. These dates are approximate and are subject to change):
 - 2.3.1 Plan submitted: September 15, 2020
 - 2.3.2 Initial item reviews of assessment materials and components: Fall 2020 – Spring 2021
 - 2.3.3 Standard setting: 2020-2021

2.3.4 Subsequent reviews: Spring 2021

2.3.5 Operational Assessments ready to administer: September 1, 2021

2.3.6 Annual Technical report: September 1, 2022

- 2.4 The supplier shall provide OEQA, within four (4) weeks of a test administration window, unduplicated reports related to tests taken, student scores by institution, institutional verification rosters, and other reports as identified by the OEQA.
- 2.5 The supplier shall provide data in a secure format after each administration that shall include, but not limited to, item-level data by candidate for all assessments.
- 2.6 The supplier shall provide an item bank report for each test after each administration year. This report shall include the total number of items, number of new items that have been piloted on the operational forms during the year, number of retired items, and the number of poor-performing items that were removed from the test bank. Noted deficiencies in the item bank shall be accompanied with a specific plan for correcting the deficiencies.

3. Criteria for General Education Test

3.1 The general education test shall meet the following criteria:

- 3.1.1 A comprehensive assessment measuring general education content knowledge and critical thinking skills of educator candidates that shall represent and assess the level of knowledge and skills required for effective educators preparing P-12 students for college and workforce readiness. The supplier shall be required to demonstrate the relationship of the test content (e.g., items and any associated rubrics) to a job analysis of beginning teaching skills.
- 3.1.2 Oklahoma educators shall be involved in the development and review of the general education test.
- 3.1.3 The test may contain a variety of item types (e.g., selected-response, constructed-response, technology enhanced); however, a constructed-response item assessing knowledge, skills, and performance is required. The content of the test shall be as authentic as possible. Meaning, all test items and scoring criteria shall be aligned to the test framework in terms of content and cognitive level. All item types shall be reliably scored and cost effective. The supplier shall specify a plan for exploring, piloting, and evaluating new item types.

- 3.1.4 There shall be a minimum of five (5) forms of the general education test that shall be created for each administration year. Each form shall be equated to a base form on which the cut score was set to either maintain the same level of difficulty (i.e., same meaning of the raw cut score) or the cut score shall be adjusted through equating to ensure the same performance expectation across test forms.

4. Criteria for Subject Matter Tests

4.1 The subject matter tests shall meet the following criteria:

- 4.1.1 The test content shall be consistent with the components of the Oklahoma Full Subject Matter Competencies for Licensure and Certification (<https://sde.ok.gov/documents/2011-12-29/full-subject-matter-competencies-licensure-and-certification>), the Oklahoma Academic Standards (<https://sde.ok.gov/oklahoma-academic-standards>), and specialized learned societies and cover the content areas listed in Appendix A.
- 4.1.2 Oklahoma educators shall be involved in the development and review of the subject matter tests to ensure alignment to the Oklahoma competencies.
- 4.1.3 The Elementary Education subject matter test shall be designed to ensure examinees have the sufficient knowledge in each content area required by the certification field in order to successfully complete the exam.
- 4.1.4 The tests may contain a variety of item types (e.g., selected-response, constructed-response, technology-enhanced); however, a constructed-response item assessing knowledge, skills, and performance is required. The content of the test shall be as authentic as possible. Meaning, all test items and scoring criteria shall be aligned to the test framework in terms of content and cognitive level. All item types shall be reliably scored and cost effective. The supplier shall specify a plan for exploring, piloting, and evaluating new item types.
- 4.1.5 There shall be multiple forms of each subject matter test created for each administration year, with the exact number of forms determined by the expected administration volume. The supplier shall identify decision guidelines for the number of forms to administer each year based on the prior year's candidate volume (e.g., volumes between X and Y shall have two forms produced for the following year). Each form shall be equated to a base form on which the cut score was set to either maintain the same level of difficulty (i.e., same meaning of the raw cut score) or the cut score shall be adjusted through equating to ensure the same performance expectation across test forms.

5. Test Identification/Development Criteria

- 5.1 The supplier shall plan to develop/redevelop seven (7) tests per year. This shall include the steps outlined in this section (i.e., developing or redeveloping the test framework, test content, pilot testing, standard setting) as well as producing a study guide and practice test for each test. The specific tests to be developed/redeveloped shall be determined based on:
 - 5.1.1 Legislation, job requirements, and state education policy in Oklahoma
 - 5.1.2 Current knowledge base for each test area and meeting the assessment objectives by the OEQA
 - 5.1.3 Results of the content validation survey reflecting a need for change
- 5.2 Test development and redevelopment activities shall include involvement by committees of Oklahoma teachers and teacher educators. These activities shall include test framework review, item bank review, and participation in standard-setting processes. The supplier shall be responsible for recruiting and organizing committees. All costs for conducting committee meetings shall be borne by the supplier. The OEQA shall be consulted about the committee(s) structure, and about the logistics (i.e., time and location) of each committee meeting. The OEQA shall be presented with a final list of committee candidates and shall take action related to endorsing the committee participants.
- 5.3 The supplier shall be responsible for developing the framework for each exam. The basis of these shall be the Oklahoma Full Subject Matter Competencies for Licensure and Certification (<https://sde.ok.gov/sites/default/files/documents/files/Competencies.pdf>), the Oklahoma Academic Standards (<https://sde.ok.gov/oklahoma-academic-standards>), and specialized learned societies.
- 5.4 Each draft framework shall be reviewed by Oklahoma educators who shall have the ability to edit the content and organization of the draft. This committee shall have the opportunity to recommend weighting of different components of the draft framework. For most of the programs, the draft framework shall be formatted into a survey and sent to Oklahoma educators for input on the importance of each component to an entry-level educator in Oklahoma. The supplier shall work with OEQA to develop a plan for distributing this survey. The results of the survey shall be consolidated and used to inform the weighting of the exam content. For the low-incidence fields (i.e., those with fewer than twenty (20) candidates per year), the supplier shall determine the weighting of the content based on input from the in-person meeting of Oklahoma educators.

- 5.5 The supplier shall make available, to OEQA for review, a prototype or shell of the stimuli and rubrics that provide the framework for producing alternative forms of the tests. Evidence of both content and statistical comparability of each form to the prototype and across forms shall be provided.
- 5.6 The supplier shall be responsible for developing all test content (e.g., selected-response items, constructed response items, scoring rubrics).
- 5.7 Any proposed scoring rubric shall include descriptions and definitions of score points that do not include words such as acceptable, adequate, satisfactory, or similar qualitative terms that imply a cut score. In the selection of marker (i.e., anchor, benchmark) papers to be used to train scorers, there shall be at least five (5) reasonable examples of scores at each score point. A reasonable example is one (1) that has received exact agreement by scorers who scored the pilot test.
- 5.8 The process of aligning test items to competencies shall incorporate two (2) levels of judgement. One (1) level is the extent that the content of the item matches the content of the competency. The second level is the extent that the cognitive complexity of the item matches the cognitive complexity of the competency. To the extent possible, an indication of breadth of coverage of each competency shall be provided (i.e., for competencies that are broadly stated, items representing all dimensions of the competency shall be developed).
- 5.9 A review of all test content shall be conducted to ensure a lack of systematic bias due to gender, race ethnicity, or any other factors not related to test content. The supplier may choose to conduct this review with Oklahoma educators after providing thorough training and examples as to how bias is identified in test content.
- 5.10 The development/redevelopment of all tests shall include, at the supplier's expense, appropriate pilot testing within the State of Oklahoma following professionally accepted standards. The primary method of pilot testing shall be through embedding items on operational forms. The data provided by embedding items shall be supplemented by stand-alone pilot sessions with examinees who are representative of the candidates who shall be taking the test. OEQA shall assist in arranging for pilot participants and test sites. For the smaller subject matter testing programs, the supplier shall work with OEQA to conduct a pilot focus group to review the draft content. The results of all pilot test activities shall be reported to OEQA along with the recommendations as to which items shall be used in operational administrations.
- 5.11 Standard setting committees consisting of Oklahoma educators shall be convened to set cut scores on each test. The supplier shall use a standard setting method approved by the OEQA and shall incorporate candidate performance data into the standard setting process.

6. Administrative Criteria

- 6.1 Suppliers may submit proposals in the following manner:
 - 6.1.1 A proposal for the general education test and/or
 - 6.1.2 A proposal for the subject matter test(s) (see Appendix A) and /or
 - 6.1.3 A proposal for the general education test and subject matter test(s).
 - 6.1.4 The general education test and the subject matter test(s) shall each be administered within a four-hour or less time frame. Once operational, the total cost of the tests shall be borne by the candidates. All test administration related costs to candidates shall be incorporated into one testing fee per test. All fees for this assessment shall be approved by the OEQA and remain the same for the duration of the contract. The supplier shall reimburse OEQA, on an annual basis, ten (10) dollars per test administered within the contract period.
- 6.2 The supplier shall make available study guides, preparatory materials, and practice tests that may be used by candidates to assist them in preparing for the test. At a minimum, the study guides shall contain the competencies to be measured, a bibliography or reference list helpful in preparing for the examinations, illustrative stimuli, and scoring criteria. Moreover, the guides shall be available online at no cost. The practice tests can be made available at a cost to candidates and reflect the same content and item types as are part of the operational tests. Candidates shall receive feedback on their performance on the practice tests.
- 6.3 The tests shall be administered a minimum of six (6) times annually; the dates shall be approved by OEQA.
- 6.4 The test shall be administered in assessment centers in multiple locations in Oklahoma and across the country. All necessary training for each testing site shall be conducted by the supplier. OEQA shall approve the testing sites. The supplier shall provide out-of-country testing services (i.e. test administration) at United States military bases. All testing centers shall supplier the same standardized conditions for taking the test including screen size and optional resources (e.g., white boards). The supplier shall include a description of the test security and monitoring processes that shall be in place at all testing centers.
- 6.5 The supplier shall be responsible for the design and production of all test forms, answer forms, registration materials and publicity materials; the registering of candidates and employing site administrators; test distribution; and test administration.

- 6.6 A secure, on-line registration option shall be available to candidates. The on-line site for registration shall include relevant information about the program in addition to the registration application.
- 6.7 Electronic distribution of registration materials is preferred.
- 6.8 The supplier shall demonstrate that the test item bank is sufficient to support multiple forms and/or administration modes of the test. Each form shall be statistically equated to the base form.
- 6.9 The supplier shall provide a customer service line for candidates and institutions for use in obtaining information about test administration. The supplier shall provide staffing for the telephone service during business hours in the central time zone. The number and hours of operation shall be published online.
- 6.10 The supplier shall provide test results in a timely manner so that candidates needing to retest can do so within the adjacent testing window.
- 6.11 The supplier shall develop all directions for administration and provide training in the administration of the test to test administrators and proctors, as necessary. The supplier shall develop administration procedures suitable for use with examinees who have special needs or who require adaptations and accommodations due to disabilities, including disabilities affecting vision and hearing.
- 6.12 The supplier shall be responsible for all scoring and score reporting.
- 6.13 Scoring verification shall be conducted, following professionally accepted standards, such that for selected-response items the scoring accuracy shall be 99.9% accurate for each item and 99 % accurate overall. Scoring accuracy for other than selected response items shall be judged based on evidence of at least a level of 80% exact agreement and 95% inter-judge exact and adjacent agreement combined, assuming a four-point scale. Supplier shall provide a description of procedures to be used if constructed-response scoring reliability requirements are not met for any given scoring cycle.
- 6.14 Score reports shall be available to candidates, to the OEQA, and to other education entities (e.g., higher education institutions, state department of education) as requested by OEQA, within four (4) weeks of each test administration date or end of test administration window. Score reports shall be in a form usable by the OEQA for making decisions about whether a candidate has met the components of the licensure requirements associated with performance on this test and to be able to integrate the scores for each candidate with other OEQA data files containing candidate information. Data shall be available to each institution of higher education that shall show the individual score reports for all students that designated that institution to receive a report.

- 6.15 Data shall also be available which shall show the institution(s) from which the candidate received the majority of the general education coursework, subject matter coursework, as well as the institution from which the candidate's degree was received. Test score data shall also be reported for each institution designated by OEQA (i.e., Title II data).
- 6.16 Prior to the release of scores, institutions shall be provided a list of examinees in order to verify their candidates.
- 6.17 A summary report following each administration of the test shall be provided that shall permit comparison of institutions based on student responses on an aggregate by:
- 6.17.1 Total score
 - 6.17.2 Sub-score level
 - 6.17.3 Competency level
- 6.18 A summary analysis of scores by test, gender, race ethnicity, documented disability, route to certification, educator preparation provider, and college major shall be provided annually.
- 6.19 Additional technical reports consistent with relevant portions of AERA, APA, NCME (2014) *Standards for Educational and Psychological Testing* shall be required. These reports shall contain information related to any developmental activities undertaken and evidence of the psychometric quality of all tests included in this testing program. The minimum evidence of psychometric quality is provided in the Technical Requirements section of the RFP. Reports shall be delivered after each administration of the assessment. An annual summary report shall also be provided.
- 6.20 The supplier shall be available to make semi-annual presentations to OEQA either in person or through electronic media, at the expense of the supplier.
- 6.21 For the purpose of legal defensibility, the supplier shall maintain complete files of all documents, minutes of meetings, and correspondence related to the development, validation, field testing, and/or administration scoring of the examination and the procedures used in validation of the assessment.
- 6.22 The supplier shall make available a plan for test site security that includes collection of examinee thumbprints.

- 6.23 Candidates who fail an exam shall have the option to request, within one (1) calendar year, a rescore for additional feedback on their test performance within a minimum four (4) week period from the time of the request. The supplier shall describe the type of feedback they shall provide to candidates and the cost to the candidate for this service.

7. Pilot Testing and Reports

7.1 Content Pilot Testing

7.1.1 For tests that are likely to be taken by more than fifty (50) candidates annually, all items, both selected and constructed response, shall be pilot tested using a sample of at least fifty (50) examinees who are representative of the candidates who shall be taking the test. The primary method of pilot testing shall be through embedding items on operational forms. The data provided by embedding items shall be supplemented by stand-alone pilot testing sessions with examinees who are representative of the candidates who shall be taking the test (i.e., same college majors).

7.1.2 For tests that are likely to be taken by fifty (50) or fewer candidates annually, all items, both selected and constructed response, shall be pilot tested with a focus group representative of the candidates who shall be taking the test. The supplier shall provide OEQA with a report of the process and findings from these focus groups to review.

7.2 Prior to the start of each administration year, the supplier shall provide OEQA with a report of the forms prepared for the administration year including:

7.2.1 The number of forms created for each test and field based on the anticipated testing volume.

7.2.2 The overlap between forms within each field (i.e., percentage selected response and constructed response)

7.3 The supplier shall be responsible for providing OEQA, at least annually, with a full technical report for the program. This shall include all test, form, and item-level statistics. The following sections detail the expected contents of this report and standards by which the technical quality shall be judged.

7.4 At the exam level (i.e. test field), the supplier shall include in the report:

7.4.1 The number of tests administered

7.4.2 Examinee performance (i.e., score mean, median, distribution, pass rate)

7.4.3 Scoring accuracy. For the selected-response sections of the exams, this includes the accuracy of the selected-response scores. For the constructed-response sections, this includes the inter-judge agreement. OEQA expects evidence of at least a level of combined (assuming a four-point scale.)

7.5 At the test form level, the supplier shall include in the report:

7.5.1 The number of tests administered.

7.5.2 Examinee performance (i.e., score mean, median, distribution, pass rate)

7.5.3 Reliability at the section (e.g., selected-response, constructed-response) and total test levels

7.5.4 Item statistical characteristics (e.g., p-value, discrimination)

7.5.5 Each form test of each test (and/or subtest), for which a passing score is set and is taken by more than fifty (50) candidates annually, shall meet the following criteria:

7.5.5.1 Total score reliability estimates for the selected and constructed response items combined, based on an acceptable index of internal consistency, shall be at least .80.

7.5.5.2 Decision reliability estimates, based on an acceptable index of decision consistency, shall be at least .90.

7.5.5.3 For any test form that does not meet the expectations for either total score reliability or decision reliability, the report shall indicate the contactor's plans to resolve the problem.

7.6 At the item level, the supplier shall include in the report:

7.6.1 Measures of item statistical performance for selected-response items. For those items that are common across forms, the supplier shall combine the information across forms to provide one overall evaluation of each unique item. This analysis shall include item-to-total correlations and item difficulties. Specific to these values, OEQA expects the supplier maintain item-to-total correlations greater than .20 and item difficulties (p-values) between .30 and .90. Items that do not meet these requirements shall be flagged for review and potential elimination from the item bank. The data provided on item performance shall be accompanied by a report that indicates any items for which the above expectations are not met. The report shall indicate the supplier's plans for these items.

7.6.2 For constructed-response items, this shall include the average score as well as the distribution of scores across the score scale.

7.7 The technical report shall provide evaluation at the exam, test form, and item/prompt levels. Specifically, the supplier shall provide an interpretation of the results identifying any areas of weakness in the exam and what changes have been/shall be made to address these areas. The technical report shall provide detailed descriptions of all updating activities that occurred since the last technical report. Updating information includes test revisions that resulted from incorporating performance items, or any other updating activities.