# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>3</td>
</tr>
<tr>
<td>Purpose of the Missouri Educator Gateway Assessments (MEGA)</td>
<td>4</td>
</tr>
<tr>
<td>Development of the Missouri General Education Assessment (MoGEA)</td>
<td>6</td>
</tr>
<tr>
<td>The Test Development Process</td>
<td>6</td>
</tr>
<tr>
<td>Task 1. Conduct Program Planning</td>
<td>6</td>
</tr>
<tr>
<td>Task 2. Establish Advisory Committees</td>
<td>7</td>
</tr>
<tr>
<td>Task 3. Prepare Assessment Framework</td>
<td>8</td>
</tr>
<tr>
<td>Task 4. Review Assessment Framework</td>
<td>11</td>
</tr>
<tr>
<td>Task 5. Prepare Assessment Items</td>
<td>12</td>
</tr>
<tr>
<td>Task 6. Review Assessment for Potential Bias</td>
<td>13</td>
</tr>
<tr>
<td>Task 7. Conduct Field Test</td>
<td>14</td>
</tr>
<tr>
<td>Task 8. Conduct Marker Response Establishment</td>
<td>15</td>
</tr>
<tr>
<td>Task 9. Develop Test Preparation Materials</td>
<td>15</td>
</tr>
<tr>
<td>Task 10. Begin Operational Assessment Administration</td>
<td>16</td>
</tr>
<tr>
<td>Task 11. Conduct Confirmatory Item Validation &amp; Standard Setting</td>
<td>16</td>
</tr>
<tr>
<td>Technical Properties of the MoGEA</td>
<td>19</td>
</tr>
<tr>
<td>Score Reporting</td>
<td>25</td>
</tr>
</tbody>
</table>
Preface

The Missouri Department of Elementary and Secondary Education (DESE) has contracted with the Evaluation Systems group of Pearson (Evaluation Systems) to plan, develop and administer the Missouri Educator Gateway Assessments (MEGA) program. The MEGA program, designed to support the state’s vision for education reform, includes the Missouri General Education Assessment (MoGEA), the Missouri Educator Profile (MEP), and 56 Content and Pedagogy assessments. All assessments are administered by computer.

Assessment development for the MEGA program was divided into two phases. The MoGEA and the MEP were developed in Phase One and became operational in September 2013. A Paraprofessional Assessment, administered by Evaluation Systems, also went operational in Missouri in September 2013. The Content and Pedagogy Assessments were developed in Phase Two, with administration beginning in September 2014.

The MoGEA was developed in 2013 and was first operational in September 2013. In the spring of 2014, DESE asked Evaluation Systems to revise the assessment. The original MoGEA has five subtests: English Language Arts, Writing, Mathematics, Science, and Social Studies. The revised MoGEA will have four subtests in Reading Comprehension and Interpretation, Writing, Mathematics, and Science and Social Studies. The explicit purpose of the MoGEA is to help identify candidates with the requisite level of knowledge and skills that Missouri educators have identified as necessary to perform successfully in educator preparation programs in Missouri. This technical manual documents the development of the original MoGEA.

This MoGEA technical report provides information to support the validity of MoGEA score interpretations. The report focuses on development and administration of the original MoGEA, which has been used as a required assessment for entry into educator preparation programs in the state of Missouri beginning in September 2013. The original MoGEA will be replaced in late August 2015, when the revised MoGEA becomes operational. This report is intended to inform policy makers, state educators and other interested stakeholders about the MoGEA, its purposes and development processes, and characteristics of data produced by the MoGEA.

The report first describes the purpose and composition of the MEGA program, followed by a description of the processes used in the development of the MoGEA, including the test framework development, item development, content and bias reviews, and establishment of the passing standard. The last two sections present technical characteristics and score reporting for the assessment.
Purpose of the Missouri Educator Gateway Assessments (MEGA)

The Missouri Department of Elementary and Secondary Education (DESE) has embarked on a “Top 10 by 20” initiative. The goal of the initiative is for Missouri to be one of the top ten states in the nation for student achievement by 2020. In 2012, DESE contracted with Evaluation Systems to develop and administer an assessment required for entry into Missouri educator preparation programs; and develop and administer new teacher and school administrator certification tests.

Pearson developed the MEGA program to help assess the basic skills and content knowledge needed by candidates entering educator preparation programs, and the content knowledge and pedagogical skills needed by teacher and public school administrator candidates for Missouri state certification, as defined by the Missouri State Board of Education. The MEGA program includes the MoGEA, an assessment required of all students entering an educator preparation program in the state; the Missouri Educator Profile (MEP), an inventory of work styles used in educator preparation programs to counsel and develop student candidates; and the Missouri Content and Pedagogy Assessments, required for certification as teachers and school and district administrators in the state. The MEGA program is aligned with Missouri and, where applicable, national standards.

The state of Missouri can be confident in the quality of all tests in the MEGA program. The assessments were developed in accordance with the practices recommended by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999 and 2014), which require a clear definition of the content domain being assessed and a rationale to support a claim that the knowledge and skills being assessed are required.

To ensure that the MoGEA reflects essential knowledge and skills for educator preparation program candidates, Evaluation Systems used widely accepted professional standards to draft the assessment framework. Practitioners at the PK-12 level and educator preparation program faculty throughout the state of Missouri were involved in reviewing the assessment materials to confirm that they capture the requisite knowledge and skills a student in an educator program must possess to be successful.

All MEGA program tests are offered via computer. The computer-based assessment is available year-round, by appointment, Monday through Saturday (excluding some holidays). The MEGA program is delivered through a national network of Pearson’s computer-based testing centers, as well as testing centers at university and other locations throughout Missouri.

The major purpose of testing is to allow stakeholders to draw inferences about individuals regarding the knowledge and skills they possess. Accurate inferences can be drawn from test scores only if there is agreement between what candidates are expected to know and what appears on a test. Therefore, it is important that a
test assess the knowledge and skills that candidates are expected to possess as outlined in the test framework that documents the content subtests/domains and associated competencies.

The MoGEA content has been fully aligned with the academic standards that the state of Missouri has deemed appropriate for the level of general education skills to perform successfully in an educator preparation program. The MoGEA was carefully developed to provide a comprehensive general education assessment based on the state’s professionally reviewed and accepted standards. Evaluation Systems worked with DESE to ensure that the content of the MoGEA is appropriate as defined by these standards.
Development of the Missouri General Education Assessment (MoGEA)

Test development involves a comprehensive process of defining and structuring a test, collecting validity evidence regarding test content, reviewing content to ensure it is equitable and free from bias, developing and piloting test items, and establishing an appropriate passing standard through guided standard setting activities. Evaluation Systems developed the MoGEA using such a process, collecting key validity evidence throughout to support the use of the test for the purpose of placing student candidates in educator preparation programs in Missouri.

The Test Development Process

Evaluation Systems developed the MoGEA in accordance with the guidelines specified in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999 and 2014) for defining test content, developing test items, establishing passing standards, and collecting evidence to support the validity of the assessment. Evaluation Systems incorporated these guidelines for test development for the MoGEA in eleven distinct steps:

- Task 1. Conduct Program Planning
- Task 2. Establish Advisory Committees
- Task 3. Prepare Assessment Framework
- Task 4. Review Assessment Framework
- Task 5. Prepare Assessment Items
- Task 6. Review Assessment Items for Potential Bias
- Task 7. Conduct Field Test
- Task 8. Conduct Marker Response Establishment
- Task 9. Develop Test Preparation Materials
- Task 10. Begin Operational Assessment Administration
- Task 11. Conduct Confirmatory Item Validation and Standard Setting

To support the connection between the MoGEA and its educational purposes, validity was a central focus in the test development process. Validity evidence for each of these steps in the development process was carefully compiled and documented. Additionally, the test development process involved careful attention to bias prevention in order to create a fair assessment. Details for each step are described below.

Task 1. Conduct Program Planning

Evaluation Systems met with representatives of DESE in Missouri and by conference calls, as needed, to discuss and formulate plans for program development activities. Topics discussed included identification of Missouri and national standards to inform
development, proposed conference dates, the recruitment of Missouri educators to participate in assessment development activities, and the involvement of Missouri stakeholder groups.

Evaluation Systems and DESE conducted weekly conference calls to monitor and review progress on assessment development.

**Task 2. Establish Advisory Committees**

DESE and Evaluation Systems worked with two primary groups of educators throughout the MoGEA development process: a Bias Review Committee and a Content Advisory Committee. Committees were composed of certified public school educators and educator preparation program faculty in Missouri as nominated by DESE, professional organizations, teacher educator programs, and school superintendents and principals. Educators applied for participation in a committee and were selected by DESE. The committee members participated in the following roles.

**Bias Review Committee**

While bias prevention is an integral part of Evaluation Systems test development activities and a component of the Content Advisory Committee’s (CAC) responsibility, Evaluation Systems established a separate and independent Bias Review Committee (BRC) in Missouri to focus primarily on reviewing assessment materials for potential bias issues, a developmental step recommended by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999 and 2014).

The BRC was composed of Missouri public school educators and preparation program faculty. Evaluation Systems convened the BRC to review the draft test frameworks and test items for fairness and sensitivity, identify any potential sources of bias based on review criteria provided, and make recommendations on ways to eliminate potential sources of bias.

Evaluation Systems provided the BRC with information regarding the background, purpose, and policies of the MEGA program and the purpose of the MoGEA. Explicit bias review criteria were presented to bias review committee members during the training orientation to each conference and through the relevant orientation manuals. Committee members reviewed the test framework and test items for bias based upon a set of review criteria relating to areas of content, language, offense, stereotypes, inclusion, and fairness.

**Content Advisory Committee**

A Content Advisory Committee (CAC) was established for the MoGEA. The CAC was composed of Missouri public school educators and preparation program faculty.
Content Advisory Committee members were selected with consideration given to the following criteria:

- Gender, racial, ethnic, and cultural diversity
- Representation across educational levels (e.g., elementary, middle, and secondary levels)
- Representation from diverse school settings (e.g., urban areas, rural areas, large schools, small schools)
- Geographic representation

The various roles of the CAC were to review the assessment framework, review marker responses for the Writing Subtest, and participate in the Item Validation/Standard Setting Conference.

**Task 3. Prepare Assessment Framework**

The first step in developing test specifications is to extend the original statement of purpose(s), and the construct or content domain being considered, into a framework for the test that describes the extent of the domain, or the scope of the construct to be measured. *Content specifications*, sometimes referred to as *content frameworks*, delineate the aspects (e.g., content, skills, processes, and diagnostic features) of the construct or domain to be measured. The delineation of the content specifications can be guided by theory or by an analysis of the content domain (e.g., an analysis of job requirements in the case of many credentialing or employment tests). The content specifications serve as a guide to subsequent test evaluation. (AERA, APA, & NCME, 2014, p. 76)

Evaluation Systems prepared a framework and test design for the MoGEA for review and approval by the Missouri DESE.

The test design delineates item formats, response formats, scoring specifications and other aspects of the assessment. The design outlines the number of multiple-choice items and constructed-response items as well as the percentage of total assessment score allocated to the multiple-choice and constructed-response sections of the assessment. The test design also specifies the number of scorable and nonscorable items per subtest.

The MoGEA was designed to accommodate the appropriate numbers, types, and distribution of test items, in accordance with the test framework. The MoGEA consists of five subtest designed to measure basic skills and content knowledge in English language arts, writing, mathematics, science, and social studies. The English Language Arts, Mathematics, Science and Social Studies subtests are each composed of 40 multiple-choice items. The Writing subtest consists of one constructed-response item.

The MoGEA has five subtests:
Subtest 1: English Language Arts  
Subtest 2: Writing  
Subtest 3: Mathematics  
Subtest 4: Science  
Subtest 5: Social Studies

The framework for MoGEA clearly defines the content knowledge and skills important for success in an educator preparation program. The *Standards for Educational and Psychological Testing* require that evidence be provided to show that the knowledge and skills that the test intends to assess are required for success in a Missouri educator preparation programs (AERA, APA, & NCME, 1999 and 2014). In developing the framework, Evaluation Systems worked with DESE to identify appropriate standards.

The MoGEA framework is based on the Missouri General Education Standards which represent the knowledge and skills described in textbooks, curriculum guides, certification standards, and the learning goals as defined by the Missouri State Legislature.

Some of the standards that were consulted during the MoGEA test development process include:

- **Missouri General Education Entry and Exit Competencies**
  - Missouri Freshman Composition Exit Sequence
  - Missouri Oral Communication/Public Speaking Exit Competencies
  - Missouri Mathematics Entry Competencies
  - Missouri College Algebra Exit Competencies
  - Missouri Entry Science Competencies
  - Missouri Introduction to Biology Exit Competencies
  - Missouri Introduction to Chemistry Exit Competencies
  - Missouri Social Sciences Entry Competencies
  - Missouri United States History Exit Competencies
  - Missouri American Government Exit Competencies

- **Common Core State Standards (CCSS)**
  - CCSS High School Algebra: Arithmetic with Polynomials and Rational Expressions
  - CCSS High School Algebra: Creating Equations
  - CCSS High School Algebra: Reasoning with Equations and Inequalities
  - CCSS High School Algebra: Seeing Structure in Expressions
  - CCSS High School Functions: Trigonometric Functions
  - CCSS High School Functions: Interpreting Functions
  - CCSS High School Functions: Linear, Quadratic, and Exponential Models
  - CCSS High School Geometry: Congruence
  - CCSS High School Geometry: Modeling with Geometry
  - CCSS High School Geometry: Geometric Measurement and Dimension
- CCSS High School Geometry: Expressing Geometric Properties with Equations
- CCSS High School Geometry: Similarity, Right Triangles, and Trigonometry
- CCSS College and Career Language Standards 11-12
- CCSS Mathematical Practice
- CCSS High School Number and Quantity: The Complex Number System
- CCSS High School Number and Quantity: Quantities
- CCSS High School Number and Quantity: The Real Number System
- CCSS College and Career Reading Standards for Literacy in History/Social Studies 11-12
- CCSS College and Career Reading Standards for Literature 11-12
- CCSS College and Career Reading Standards for Informational Text 11-12
- CCSS High School Statistics and Probability: Making Inferences and Justifying Conclusions
- CCSS High School Statistics and Probability: Interpreting Categorical and Quantitative Data
- CCSS College and Career Speaking and Listening Standards 11-12
- CCSS College and Career Readiness Standards for Writing 11-12

  - National Standards
    - National Curriculum Standards for Social Studies
    - National Council of Teachers of Mathematics (NCTM) Principles and Standards for School Mathematics
    - International Reading Association (IRA)/National Council of Teachers of English (NCTE) Standards for English Language Arts

Evaluation Systems structured the assessment framework to include the following components:

  - **Content Subtests:** the primary areas of content knowledge to be assessed. The content subtests serve to structure the content for both test preparation and score reporting purposes.

  - **Competencies:** broad, meaningful statements of knowledge and/or skills important for performing in educator preparation programs. Collectively, the competencies define the range of content to be measured by the test.

  - **Descriptive Statements:** define each test competency. Each descriptive statement provides more detailed information about a competency and provides specific examples of the knowledge and skills eligible for testing. There are multiple descriptive statements for each competency.
Evaluation Systems organized preliminary reviews of the test framework by content experts. These reviews elicited focused feedback regarding the content and structure of the draft test framework. Finally, curriculum specialist from the Missouri DESE reviewed the framework. Both reviewer comments and DESE feedback informed development of the draft framework.

**Task 4. Review Assessment Framework**

Evaluation Systems prepared a draft assessment framework and convened the BRC to review the framework. At the meeting, members of the BRC were trained according to specified bias review criteria and then participated in a review of the framework.

BRC members used the following criteria when conducting the review of the framework.

**Content**

Does any element of the competencies or descriptive statements contain content that disadvantages a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

**Language**

Does the language used to describe any element of the competencies or descriptive statements disadvantage a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

**Offense**

Is any element of the competencies or descriptive statements presented in such a way as to offend a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

**Stereotypes**

Does any element of the competencies or descriptive statements contain language or content that reflects a stereotypical view of a group based on gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

**Fairness**

Taken as a whole, is the list of competencies and descriptive statements fair to all individuals regardless of race, gender, cultural background, or other personal characteristics?
**Diversity**

Does the list of competencies and descriptive statements permit appropriate inclusion of content that reflects the diversity of the Missouri population. Bias-related recommendations made by the BRC were incorporated into the assessment framework.

Following the BRC meeting, members of the MoGEA Content Advisory Committee (CAC) conducted a review of the framework. Content-related validity evidence provides evidence that the test is measuring the content it was designed to assess. The CAC review is an important step in establishing appropriate validity evidence for the test (AERA, APA, & NCME, 1999 and 2014). The CAC members conducted their review of the draft framework online. Committee members reviewed the test framework, including the competencies and descriptive statements based on the following criteria:

**Alignment**

Is the assessment framework consistent with the Missouri general education competencies for college undergraduates and the knowledge and skills that candidates should possess when entering an educator preparation program?

**Completeness**

Does the assessment framework address important areas of general education knowledge and skills in Missouri?

**Language and Terminology**

Does the assessment framework contain language that is stated clearly, using terminology that reflects educational policies and practices in Missouri?

**Freedom from Bias**

Is the assessment framework free from elements that might potentially disadvantage an individual because of her or his gender, sexual orientation, race, nationality, ethnicity, religion, age, disability, or cultural, economic, or geographic background?

The committee reviewed the framework and offered suggested revisions electronically. Following the review of the test framework, the framework was revised based on both the BRC and CAC’s recommendations. The framework for the Missouri General Education Assessment is found in Appendix A.

**Task 5. Prepare Assessment Items**

Evaluation Systems developed MoGEA items in accordance with the final, validated test framework and the test design. Pearson assembled a team of content specialists, test development specialists, editors, content reviewers, and equity advisors to develop the test items and associated scoring rubric to ensure that the
test materials produced were closely linked to the test framework and met Evaluation Systems’ standards for editorial quality.

Evaluation Systems prepared test items for the MoGEA by drawing from existing item banks and by drafting additional items as necessary. Test item and material development involves a series of activities designed to produce a test that is technically sound, reliable, and valid. Evaluation Systems used the following two item formats.

- Multiple-choice items, which are intended to address specific knowledge, skills, and abilities in a rigorous, authentic, and challenging manner
- Constructed-response items, which require candidates to provide a written response to an assignment

Task 6. Review Assessment Items for Potential Bias

Evaluation Systems conducted an item review with the Bias Review Committee (BRC). The BRC was convened to review test items for potential bias in relation to established review criteria. Committee members were oriented to the background, purposes, and policies of the MEGA program and the purpose of the MoGEA. Pearson presented review criteria in the training orientation and in the orientation manual. Committee members reviewed the test items for bias based upon the following review criteria.

Content
Does the item contain content that disadvantages a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

Language
Does the item contain language that disadvantages a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

Offense
Is the item presented in such a way as to offend a person because of her or his gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?

Stereotypes
Does the item contain language or content that reflects a stereotypical view of a group based on gender, race, nationality, ethnicity, sexual orientation, religion, age, disability, or cultural, economic, or geographic background?
**Fairness**

Taken as a whole, are the items fair to all individuals regardless of race, gender, cultural background, or other personal characteristics?

**Diversity**

Taken as a whole, do the items permit appropriate inclusion of content that reflects the diversity of the Missouri population?

The committee either approved the assessment items as presented, made recommendations for revision, or deleted assessment items from the item bank.

**Task 7. Conduct Field Test**

Field testing provides another source of validity evidence by gathering data regarding the performance characteristics of the draft test items. Item performance can be assessed and, when needed, items can be revised for future pilot tests and administration.

Evaluation Systems conducted a state-wide field test of the MoGEA Writing Subtest items to obtain sample responses and data about the items’ statistical and qualitative characteristics. Adopted multiple-choice items were not field tested because items used in other programs appeared on the first operational test forms. Additional multiple-choice items are introduced as nonscorable items on operational test forms.

The field test for the Writing constructed-response items was designed to determine how items perform, identify any items that may need revision, and to evaluate the comparability of items within a set. Constructed-response items should be comparable to each other in difficulty and other relevant characteristics. Field testing of the Writing Subtest constructed-response items also provided an opportunity to gather candidate responses that could be used as marker or exemplar responses for scorer training or study guide sample responses.

Evaluation Systems administered field tests at institutions in the state of Missouri offering state-approved educator preparation programs. Field tests were administered under conditions of quiet, confidentiality, and test security, and were not permitted access to unauthorized aids, such as reference materials, so as to approximate operational testing conditions.

Eligible participants for the MoGEA field test included candidates that were enrolled in Missouri educator preparation programs and candidates anticipating enrolling in a
Missouri educator preparation program. Evaluation Systems recruited field test participants through notifications to deans and other college contacts. Participants were required to indicate their eligibility for the field test and their willingness to abide by the security and confidentiality requirements of field testing by indicating their agreement with a non-disclosure statement. Multiple pilot test forms were constructed to include all constructed-response items. Pilot test participants were asked to respond to two constructed-response items.

Constructed-response items were scored using procedures and scorers in a process designed to simulate the scoring of responses for operational test administrations. Each field test response was individually scored by trained and calibrated scorers using the MoGEA scoring rubric. The scoring rubric contained performance characteristics (descriptions of characteristics of the response) to guide the scoring of constructed-response items.

Evaluation Systems generated constructed-response items statistics for review and determined if the items could be used on test forms.

**Task 8. Conduct Marker Response Establishment**

Following the field test, Evaluation Systems oversaw marker selection for the Writing Subtest constructed-response item. The CAC was convened to identify marker responses that exemplify the components of the four-point scale. At the marker response selection meeting, committee members reviewed scoring criteria and selected a set of “marker responses” that illustrate the holistic score points on the scoring scale. These marker responses are used to train scorers in evaluating operational responses to the constructed-response assignment. The use of the marker responses, together with the standardized scoring scale and performance characteristics, promotes continuity and consistency in scoring over time, across test forms, test administrations, and scorers. The marker responses also help to ensure that scores retain a consistent meaning over time, and that candidates’ responses are judged similarly regardless of when they take the test or which test form they take.

**Task 9. Develop Test Preparation Materials**

Evaluation Systems developed web-based resources to assist candidates in preparing for the MoGEA. Candidates have access to the test framework, a study guide, a tutorial on computer-based testing, and a practice test. The following describes several resources available to MoGEA candidates:

- **MoGEA Framework** – contains the test competencies organized by content domains, groupings of test competencies that reflect the main areas of subject-matter knowledge for each subtests
• **MoGEA Study Guide** – provides sample questions and general information about the test

• **Computer-Based Testing Tutorial** – familiarizes candidates with the navigational tools and operations of computer-based testing. The tutorial directs candidates on how to record, change, and review answers during the testing experience. The tutorial also gives candidates the opportunity to use functions to view visuals and exhibits, scroll pages, and review items.

• **Test Preparation Video** – provides a general overview of test preparation and test-taking strategies. It is designed to help the candidate prepare for the MoGEA by describing how to use the study resources provided online.

• **Practice Test** – replicates the computer-based testing experience by providing a full length computer-based test, allowing candidates to understand the complexity of questions designed to measure content knowledge and skills and practice using the functions of the testing platform.

### Task 10. Begin Operational Assessment Administration

Items that met Evaluation Systems’ requirements in terms of the review criteria were assembled into test forms for operational administrations. Evaluation Systems’ test construction group of experts assembled each form to meet the appropriate assessment specifications. The experts considered difficulty of the items and the ability of items to distinguish between weak and strong candidates so that overall test statistics are comparable across forms. The test construction experts also checked for inclusion of unique item types, if applicable, and verified the answer key. After the forms were constructed, independent content experts were brought in to review the assembled forms for item accuracy, topicality, overlap, and possible cueing.

The MoGEA began operational test administration in September 2013.

### Task 11. Conduct Confirmatory Item Validation and Standard Setting

A panel composed of Missouri public school educators and preparation program faculty was convened for a confirmatory item validation and standard setting conference. The *Standards for Educational and Psychological Testing* provided the foundations for the item validation and standard setting process (AERA, APA, & NCME, 1999 and 2014). The panel included both educators who previously served on the CAC and additional Missouri educators. A total of 15 educators participated in two MoGEA item validation and standard setting activities. Overall, approximately 66% of the panel members were female and 33% were male. The majority of the panel members (approximately 60%) were educator preparation program faculty, while approximately 40% were public school educators. All geographic regions of the state of Missouri were represented.
There were two activities associated with this task. During the confirmatory item validation activity the committee reviewed each item using the four criteria: match to competency, accuracy, freedom from bias, and job-relatedness. CAC members confirmed that each assessment item is valid for use on the assessment according to these criteria. Items determined by the CAC to be not valid were removed from the item bank.

During the standard setting phase of the conference, CAC members engaged in a series of activities. The first of these activities required participants to “take the test” without the benefit of the answer key, in order to simulate the candidate testing experience. The CAC members then made a series of judgments regarding each of the multiple-choice and constructed-response items.

**Multiple-Choice Items**
The standard setting process utilized a method commonly referred to as the “Modified Angoff procedure” to obtain item-by-item judgments from each panelist about the percentage of acceptable candidates who would correctly answer each item on the assessment form. Panelists made two rounds of standard setting ratings. Following the first round of ratings, panelists were provided with the results of their own and fellow panelists’ first round item judgments. Panelists also received a table showing the percentage of candidates who had answered the items correctly, for those test items that had been used in Evaluation Systems programs in other states. Panelists then completed a second round of judgments in which they had an opportunity to review and revise their Round 1 item judgments in light of the information provided.

Following the second round of ratings, the item-by-item judgments were combined to calculate a panel-based passing score.

**Constructed-Response Item**
For the constructed-response item in the Writing Subtest, panelists reviewed sample responses and descriptions of performance at each score point on the scoring scale. Panelists then identified the candidate response that best represents the performance of acceptable candidates. As with the multiple-choice items, following their first round of judgments, panelists were provided with a summary of their own and fellow panelists’ first round ratings, and used that information to provide a final recommendation in the second round.

**Standard Setting Results**
Missouri educators provided a recommendation of the level of performance deemed acceptable on each MoGEA subtest for candidates entering an educator preparation program in the state of Missouri. These recommendations were provided to DESE and formed the basis for the MoGEA score scale, which, as approved by the State Board of Education, used the panel-based judgments to define the scores reported to candidates and educator preparation programs. For each subtest, candidates who achieve the panel-based raw passing score receive a scaled score of 220 on the 100-to-300 point MEGA score scale.
For the 2013-14 and 2014-15 program years, the Board determined that individual educator preparation programs would establish their own passing standards. Each EPP then defined the score on the MEGA scale that would be considered acceptable for candidates entering their own programs.

The Standard Setting Conference Survey used to evaluate the panels understanding of the standard setting process is located in Appendix B, and the Standard Setting Conference Survey Results for the two conferences is located in Appendix C.
Technical Properties of the MoGEA

The Standards for Educational and Psychological Testing require that testing agencies provide relevant technical information about assessments so test users and reviewers have sufficient information to make judgments about the quality of the test, the resulting scores, and the interpretations based on test scores (AERA, APA, & NCME, 1999 and 2014). This information can ultimately assist test users and reviewers in determining the appropriateness of the test for its intended purpose (AERA, APA, & NCME, 1999 and 2014). The following sections outline the efforts made to ensure the quality of the MoGEA.

Scoring

The MoGEA consist of multiple-choice items and one constructed-response item. The scoring procedures are carefully documented for the multiple-choice and constructed-response items. Additionally, the reliability of holistic scoring for the constructed-response item is monitored through multiple analyses of scorer performance.

The correct responses for multiple-choice items are an integral component of each item definition, so that the selection of test items that appear on a test form automatically results in the generation of answer keys during the test construction process. These keys are reviewed and checked at several points during development and test form production. 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 the MoGEA English Language Arts, Mathematics, Science, and Social Studies subtests, each of which is composed of 32 scorable multiple-choice items, the final raw score is the total number of correct responses on the test. These raw scores are transformed and reported on a scale ranging from 100 to 300, with 220 representing the Board-approved state benchmark.

The Writing subtest is composed of one constructed-response item. Candidate responses to the Writing constructed-response item are scored using a focused holistic scoring methodology. In this method, scorers judge the overall effectiveness of each response using a set of performance characteristics.

Candidate responses are scored on a scale of 1 to 4, where 1 represents little or no command of the performance characteristics, and 4 represents a strong command of the characteristics. Each constructed-response item is independently scored by two scorers, and these scores are summed for a total possible score range of 2 to 8.
To participate in the holistic scoring process, scorers must meet qualifications, including holding educator certification or having experience as a college educator responsible for preparing perspective teachers. Prior to scoring, each scorer receives orientation and training by a Chief Reader. The Chief Reader is responsible for training each scorer to holistically evaluate the constructed-response item. Scorers are provided with the performance characteristics and score scale for the assignment.

The Chief Reader leads training sessions in order to calibrate scorers, and monitors the scoring session to ensure that every constructed-response is scored accurately. During the training sessions, scorers practice using scoring training sets of responses to which scores have already been assigned, including historical anchor responses. The training emphasizes equity and fair application of the score scale. Once scorers have been trained, they have to prove their ability to score accurately by completing a calibration exercise.

Evaluation Systems monitors the performance of scorers throughout the focused holistic scoring process. Specific areas monitored include scorers’ ability to understand and apply the established score scale, the consistency of the scores assigned in comparison with those scored by a second scorer, and the scorers’ consistency over time. Scorers must demonstrate continued scoring accuracy.

**Item Analyses**

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

For multiple-choice items, data are collected for 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 reviewed for the current content administration and cumulatively (i.e., 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 candidates selecting each response category);
- Item-to-test point biserial correlation; and
- Mean score by response choice (average score on the total multiple-choice set achieved by all candidates selecting each response option).

Those items that do not perform within defined statistical parameters are flagged for additional review.
**Test Equating**

Each MoGEA subtests consist of multiple forms. Multiple forms are utilized across test administrations to address issues of item exposure and security. Statistical adjustment (equating) is implemented to adjust for small differences in difficulty across forms.

According to the *Standards for Educational and Psychological Testing*, equating refers to the process of placing scores from alternate, parallel, or equivalent forms on a test on a common scale (AERA, APA, & NCME, 1999 and 2014). The central purpose of the statistical equating method is to compensate statistically for possible variability in the characteristics of test forms that may affect candidate scores (e.g., differences in the overall difficulty of a new test form compared to a previous test form). Statistical equating ensures that a candidate’s scaled score is adjusted for the relative difficulty of a particular test form. Statistical equating allows test developers to attribute differences in scores across test forms to differences in the knowledge and skills of candidates, and not differences in the tests.

**Scaled Scores**

The *Standards for Educational and Psychological Testing* state that scaled scores may aid in interpretation of the test. Scaled scores allow scores to be easily comparable regardless of test form or administration (AERA, APA, & NCME, 1999 and 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 administrations and forms.

Raw test scores, number correct for multiple-choice items and holistic scores for the constructed-response item are transformed to a standard scale ranging from 100 to 300. For each MoGEA subtest, a score of 220 represents the statewide passing standard. In the case of MoGEA, the Board of Education has permitted the educator preparation programs to establish their own passing standards for the 2013/2014 and 2014/2015 academic years, so the scaled score of 220 is referred to as the state benchmark.

For the MoGEA subtests with 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 state benchmark achieve a scale score of 220, while those who achieve the maximum possible raw score will receive a scaled score of 300.
**Test Validity**

The majority of test development tasks are designed to establish and/or support the connection between the test and its educational purpose. This connection describes validity, which is the central concern in high-stakes professional testing.

*Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations. (AERA, APA, & NCME, 2014, p. 11)*

The *Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999 and 2014)* provide professional guidelines for accumulating validity evidence. The guidelines are clear that the process for accumulating such validity evidence is not a matter of performing one step during test development or gathering one set of data from time to time for statistical analysis. Rather, validation is a logical and intellectual process that begins with defining assessment content and continues during both the development and administration of the test, and is revisited through subsequent validation reviews as the test is updated. Because the collection of validity evidence is an ongoing process, Evaluation Systems gathers this evidence at every step of test development.

The validation process used by Evaluation Systems for development of the MoGEA followed professionally accepted procedures. The validation process focused primarily on establishing that the content of the assessment was appropriate for the purpose of the test. In addition, Evaluation Systems provides guidance to test takers, educator preparation programs, and statewide stakeholders regarding the appropriate interpretation and use of test scores.

Throughout the various steps in preparing the MoGEA for use, Evaluation Systems aimed to establish the validity of the assessment as recommended by the *Standards for Educational and Psychological Testing (1999 and 2014)*. The steps included:

- **Establishing the Basis for the Assessment**
  
The purpose of the MoGEA is to support entry into Missouri educator preparation programs as established by state rules and regulations.

- **Defining the test competencies**
  
The assessment competencies describe the content knowledge that a candidate should possess to succeed in an educator preparation program.
• **Conduct Content Validation of the Assessment Competencies**

Content validation of the assessment competencies occurred through alignment with documentation of content requirements and a review by Missouri DESE curriculum specialists and Missouri educators.

1. Assessment competencies were aligned with relevant laws and regulations and Missouri and national standards to provide documentation of the basis of the assessment competencies. Thus, the content of the assessment was verified as being relevant. The content alignment table for the MoGEA can be found in the Faculty Resources section of the MEGA website.

2. DESE curriculum specialists, Missouri PK-12 educators, and faculty at educator preparation programs reviewed and revised the MoGEA competencies to ensure that the alignment with Missouri and national standards were clear.

• **Validating Test Items**

A Content Advisory Committee (CAC) composed of Missouri educator preparation program faculty and PK-12 educators reviewed and validated every test item that is included in the MoGEA item bank. Educators rated an item valid if it matched the assessment competency to which it was written, was accurate, free from bias, and job-related.

• **Preventing Bias**

The prevention of bias in a testing program is both a matter of fairness and an aspect of test validity. To create a sensitive, fair, and valid assessment for test takers, Evaluation Systems makes bias prevention and equity a priority during the development and review of all tests. The Evaluation Systems manual titled *Fairness and Diversity in Testing* guides these efforts for all Evaluation System test development staff.

The manual contains four major sections that provide a discussion of the dimensions of bias in test development. The sources of bias discussed include: bias due to content, bias in language, bias due to assumptions and stereotypes, and bias due to lack of inclusion of test content that reflects diversity of the population for whom the test is intended. For each source, examples are provided to assist reviewers.

The manual also addresses specific bias prevention steps to be taken in test development and methods of bias review to apply during test development. Bias prevention steps are taken during the development of test frameworks and individual test items. The *Fairness* manual presents a comprehensive understanding of bias, including both bias prevention and equity inclusion (i.e., the inclusion of content that reflects diverse populations).
Evaluation Systems test developers and editors are charged with detecting and removing potentially biasing content, situations, language, and stereotypes. Other test development steps also reflect attention to bias prevention. The content advisory and bias advisory committee composition reflects, to the extent possible, representative demographic characteristics, as does the drawing of samples for pilot testing. Additionally, statistical analyses are designed to detect instances where one group of candidates performs much better on an item than another group of equivalent ability (Differential Item Functioning). Based on the statistics, the items can be flagged for further review.

Rather than waiting for such statistical tests to be performed, however, Pearson procedures involve the review of all test items during the initial development phase by both Pearson staff and the Missouri PK-12 educators and educator preparation program faculty who serve on the Bias Review Committee and the Content Advisory Committee.

- **Setting Passing Standards**

In addition to reviewing the framework and validating test items, Missouri PK-12 educators and educator preparation program faculty assisted in recommending the statewide benchmark scores for the MoGEA subtests. The standard setting process is described in more detail earlier in this report.

- **Communicating Appropriate Interpretations with Assessment Users**

It is important that test scores are understood and used appropriately by the various potential users of the assessment results. Evaluation Systems includes an explanatory page of text with every MoGEA examinee score report describing the included information. This information is also posted on the testing program web site. In addition, Evaluation Systems has worked closely with the state to provide guidance regarding the appropriate and psychometrically sound uses of test scores.

**Test Reliability**

The *Standards for Educational and Psychological Testing* refer to reliability as the consistency of scores when testing is repeated across administrations (AERA, APA, & NCME, 1999 and 2014). There are many common reasons for scores to fluctuate over time. Ideally, score fluctuations caused by differences in the test itself are minimized. Thus changes of test scores over time may be attributed to the candidate. Evaluation Systems uses a number of statistics to estimate test reliability for the Missouri General Education Assessment (MoGEA).
Score Reporting

After each MoGEA administration, score reports are provided to candidates to inform them of their performance on each subtest. Score reports are also provided to the Missouri DESE, and educator preparation programs, as designated by the candidate.

All candidates that register for the MoGEA online may request that a score report be e-mailed to them on the score report date published on the MEGA website. Official score reports are posted to the candidate’s online account on the score report date associated with their testing window.

The reports remain in the candidate’s online account for 45 days, during which time the candidate may view, print, and save for his or her records. After 45 days, candidates may request a copy of their scores through their online accounts on the MEGA website.

Candidate score reports include the following information:

- The date the candidate took the test
- The candidate’s scaled score, based on the number of items answered correctly converted to a scale ranging from 100 to 300, for each of the subtests taken
- Details of the candidate’s performance on each competency area assessed
- A list of educator preparation program(s), up to three, to which the candidate’s scores have been released
- The candidates testing history for the MoGEA, listing the highest scaled score obtained across all attempts for each subtest tested

Score reports are accompanied by an interpretative guide to help candidates understand the report.

Test results are sent to DESE and to educator preparation programs as designated by the candidate. These reports are delivered electronically through Evaluation System’s web-based system called ResultsAnalyzer. This interactive, electronic system is used by DESE and educator preparation programs to view, analyze, customize, download, and print reports based on operational data. The program allows users to generate reports based on administration dates, gender, race, and other demographic characteristics. Specifically, the web-based system allows users to access candidate, subtest, and program data, giving them the capability to:

- Create reports on candidate and institution results
- Customize data queries
- Aggregate data across testing program years
- Export data to other software and print graphics
- Analyze data for numerous variables
MoGEA Technical Report

This technical report provides statistical information and reports on candidate performance. Evaluation Systems prepares reports to provide information regarding the statistical characteristics of each MoGEA test form. Statistical reports contain information regarding the reliability of the test scores. As defined earlier, score reliability refers to the consistency of scores across test administrations.

There are a number of statistics that may be used to estimate test reliability. In general, reported reliability values range from zero to one, with higher values indicating greater reliability of test scores. Reliability is a property of test scores for a particular group of candidates, not a fixed property of a test. In a certification context, reliability measures may be influenced by many factors, such as:

- **Number of Candidates**

  In general, reliability estimates based on larger numbers of candidates are more stable than estimates based on smaller numbers of candidates. For this reason, reliability estimates are calculated for tests that are taken by 100 or more candidates.

- **Self-Selection of Candidates by Test Administration Period**

  Typically, examinees can decide when to take a particular test. The MoGEA is offered on computer year round. This self-selection can affect the composition, ability level, and variability of the group taking a particular test during a given testing period.

- **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 level of knowledge and skills of the candidates in the particular group taking a test on a particular occasion), the greater the reliability. Reliability estimates tend to be higher if candidates in the group have widely varying levels of knowledge and skills. Conversely, if the examinees on a particular occasion have generally similar levels of knowledge and skills, statistical estimates of reliability may tend to be lower.

- **Test Length**

  Reliability estimates tend to be higher for tests with greater numbers of test items. One obtains a more reliable estimate of a person’s knowledge 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 certification) that cover a broad range of content. Tests for educator certification typically test a broad base of knowledge and skills that pertain to certifications that will apply in a wide range of educational settings, grade levels, and educator assignments.

Several measures are employed for the MEGA program to assess the reliability of the MoGEA. Each statistical procedure employed provides different information about the reliability of the MoGEA. Measures are reported for the total test, the multiple-choice section, and the constructed-response section. The measures used on the MoGEA are described below.

**Kuder-Richardson Formula 20 (KR20) for Selected-Response Items**

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 tests composed of multiple-choice items. KR20 is reported in the range of 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.

**Scorer Agreement and Generalizability (G) Coefficient for Constructed-Response Assignments**

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 Writing Subtest Scorer Reliability Report provided in Appendix E provides selected statistics for the MoGEA Writing subtest consisting of one constructed-response item. This report includes the following scorer agreement information:

- Percent Agreement. Overall agreement determined by summing exact and adjacent agreement.
- Percent Exact.
- Percent Adjacent.
- 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 Generalizability coefficient is reported for the open-response section of each test form with at least 100 attempts. The 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 (Shavelson and Webb 1991). It is reported in the range .00 to 1.00, with a higher number indicating a greater level of generalizability.

**Total Test Decision Consistency**

There are a number of statistics that may be used to estimate test reliability. For the MoGEA, the most important testing outcome is the pass/fail decision. Total test decision consistency is a reliability statistic that describes the consistency of the pass/fail decision. For the purposes of this analysis, the statewide MoGEA benchmark is used to define the passing score. An estimate of total test decision consistency (Breyer and Lewis, 1994) is reported in the range 0.00 to 1.00; the closer the estimate is to 1.00, the more consistent (i.e., reliable) the decision.

**Test Form Statistics Report**

Evaluation Systems prepares reports to provide information regarding statistical characteristics of each test form. The MoGEA Test Form Statistics Report provided in Appendix D includes the following information, where applicable:

- Test Field Code and Name
- Form Designation
- Number of Tests Taken
- Mean
- Standard Error of Measurement (SEM)
- Decision Consistency (See the description above.)
- Stratified Alpha (Not applicable: this statistic is only for tests with both multiple-choice and constructed-response sections, so this statistic is not applicable to the MoGEA subtests)
- Test Length
- The Kuder-Richardson formula 20 (KR-20) (See the description above)
- Generalizability Coefficient (Applicable to the Writing subtest only. See the description above.)

**Total Scaled Score Distribution by Subtest**

Evaluation Systems prepares reports that display the performance of candidates on the test forms administered during the reporting period. The Total Scaled Score Distribution by MoGEA subtests shown in Appendix F provides information about the scaled score distributions for the 2013/14 academic year and for September 1, 2014 through April 12, 2015. For the MoGEA, results are reported on a scale ranging from 100 to 300. A scaled score of 220 represents the panel-based passing standard for each subtest.
The scaled score distribution includes the following information:

- **Total Scaled Score**: the inclusive list of observed total test scaled scores, in intervals of ten scale-score points
- **N**: the number of scores observed within intervals of ten scale score points
- **N at or above**: the number of scores observed at or above each scale score interval
- **Percent**: the percent of scores observed within intervals of the scaled score points
- **Percent at or above**: the percent of scores observed at or above each scaled score interval