

Board of Education Agenda Item

Item: _____ M. _____

Date: January 14, 2010

Topic: Final Review of the Advisory Board on Teacher Education and Licensure's Recommended Passing Score for the *School Leaders Licensure Assessment (SLLA)*

Presenter: Mrs. Patty S. Pitts, Assistant Superintendent for Teacher Education and Licensure

Telephone Number: (804) 371-2522

E-Mail Address: Patty.Pitts@doe.virginia.gov

Origin:

Topic presented for information only (no board action required)

Board review required by

State or federal law or regulation

Board of Education regulation

Other: _____

Action requested at this meeting Action requested at future meeting: _____ (date)

Previous Review/Action:

No previous board review/action

Previous review/action

date November 17, 2009

action First Review of the Advisory Board on Teacher Education and Licensure's Recommended Passing Score for the *School Leaders Licensure Assessment (SLLA)*

Background Information:

The 2004 General Assembly approved *House Bill 573*, [Section 22.1-298, *Code of Virginia*] that stated, in part, the following:

D. ... the Board's licensure regulations shall also require that, on and after July 1, 2005, initial licensure for principals and other school leaders, as may be determined by the Board, be contingent upon passage of the *School Leaders Licensure Assessment*.

On November 17, 2004, the Board of Education approved a passing score of 165 for the *School Leaders Licensure Assessment (SLLA)* as a requirement for all individuals seeking an initial administration and supervision endorsement authorizing them to serve as principals and assistant principals in the public schools. The effective date for implementing the passing score was July 1, 2005.

Section 22.1-298.1(E) of the *Code of Virginia* states, “The Board's regulations shall require that initial licensure for principals and assistant principals be contingent upon passage of an assessment as prescribed by the Board.” Section 8VAC20-22-590 of the *Licensure Regulations for School Personnel (effective September 21, 2007)* states that individuals seeking endorsement in administration and supervision preK-12 through the approved program route as well as an alternate route satisfy the requirements for the school leaders licensure assessment prescribed by the Board of Education. Individuals seeking an initial administration endorsement who are interested in serving as central office instructional personnel are not required to take and pass the school leaders assessment prescribed by the Board of Education. The *School Leaders Licensure Assessment* administered by Educational Testing Service is the prescribed assessment.

As part of the test regeneration process, the Educational Testing Service has completed a major revision of the *SLLA*. The changes to the assessment were significant and required completion of a standard setting study and the approval of a passing score for the revised assessment.

Although the revised *SLLA* was administered in other states beginning in September 2009, the implementation was delayed in Virginia to allow sufficient time for a state-specific standard setting study and the setting of a passing score for the assessment. A special administration of the former version of the test was held on Saturday, October 17, 2009, to allow Virginia candidates one final time to take this version. Administration of the revised *SLLA* will begin in Virginia in January 2010.

Summary of Major Elements

A Virginia standard setting study was conducted on March 24 and 25, 2009, for the revised *SLLA*. The Educational Testing Service (ETS) conducted the standard setting study on behalf of the Virginia Department of Education (VDOE) for the *SLLA*, which will be administered in Virginia for the first time in January 2010. A detailed summary of the study -- *Standard Setting Report-- School Leaders Licensure Assessment (SLLA) -- March 24-25, 2009 -- Richmond, Virginia*, is attached (Appendix A) and includes information regarding participants, methodology, and recommendations.

In addition to the Virginia specific study, the Educational Testing Service also conducted two multistate standard setting studies on April 21 and 22, 2009, in Baltimore, Maryland, and on May 12 and 13, 2009, in St. Louis, Missouri. Seventeen states participated in the two panels. On Panel I, 23 school leaders and college faculty from 16 states participated. On Panel II, 23 school leaders and college faculty participated from 15 states. The results of these studies, including the passing scores recommended by the multistate panels, are attached (Appendix B).

The purposes of the studies were to (a) recommend the minimum *SLLA* score judged necessary to award the endorsement in administration and supervision and (b) confirm the importance of the *SLLA* content specifications for entry-level school leaders in Virginia.

The revised assessment is designed to measure whether entry-level school leaders have the knowledge believed necessary for competent professional practice. The content of the assessment was defined by a National Advisory Committee of expert practitioners and preparation faculty and confirmed by a national survey of the field. The content of the revised assessment is aligned with the *Educational Leadership Policy Standards: ISSLC 2008*.

The four-hour assessment is divided into two separately timed sections:

- Section I (2 hours 20 minutes) – 100 multiple choice questions (80 operational and 20 pre-test); and
- Section II (1 hour 40 minutes) – Seven constructed-response questions calling for written answers based on scenarios and sets of documents that an education leader might encounter. Candidates are required to analyze situations and data, to propose appropriate courses of action, and to provide rationales for their proposal.

Candidate scores on the two sections are weighted such that Section I contributes 70 percent of the overall *SLLA* score and Section II contributes 30 percent. The total number of raw points that may be earned on the *SLLA* is 114 (80 points from the multiple choice section and approximately 34 points from the constructed-response section). The reporting scale for the *SLLA* ranges from 100 to 200 scaled points. A detailed description of the test is provided in the *Test at a Glance* document in Appendix C.

Prospective school leaders will be required to pay a fee for test administration and reporting results to the Virginia Department of Education. The cost for the assessment has been reduced from \$480 to \$375, including a \$50 nonrefundable registration fee.

The process used in the Virginia standard setting study is detailed in Appendix A. The panel recommended a cut score of 67.24. The next highest whole number is 68 and is considered the functional recommended cut score. The value of 68 represents approximately 60 percent of the total available 114 raw points that could be earned on the *SLLA*. The scaled score associated with 68 raw points is 154.

A similar process was used in the multistate standard setting studies as described in Appendix B. The cut score recommendations for the *SLLA* were 74.41 for Panel I and 76.16 for Panel II. These numbers were also rounded to the next highest whole number to determine the functional recommended cut scores of 75 for Panel I and 77 for Panel II. The values of 75 and 77 represent approximately 66 percent and 68 percent, respectively, of the total available 114 raw points that could be earned on the *SLLA*. The scaled scores associated with 75 and 77 raw scores are 162 and 164, respectively.

The recommended cut scores and Standard Error of Measurement (SEM) from the Virginia standard setting study and the two multistate studies are provided on the following page. The SEM is a statistical phenomenon and is unrelated to the accuracy of scoring. All test results are subject to the standard error of measurement. If a test taker were to take the same test repeatedly, with no change in his level of knowledge and preparation, it is possible that some of the resulting scores would be slightly higher or lower than the score that precisely reflects the test takers actual level of knowledge and ability. The difference between a test taker's actual score or his highest or lowest hypothetical score is known as the Standard Error of Measurement.

The Standard Error of Measurement for the recommended cut scores for the Virginia Standard Setting

Study and the multistate studies are shown below. [Note: Consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.]

Cut Scores within 1 and 2 SEMs of the Recommended Cut Score – Virginia Study

Recommended Cut Score	68	Scale Score Equivalent	154
-2 SEMs	58		143
-1 SEM	63		149
+1SEM	74		161
+2 SEMs	79		167

Cut Scores within 1 and 2 SEMs of the Recommended Cut Score – Multistate Study (Panel I)

Recommended Cut Score	75	Scale Score Equivalent	162
-2 SEMs	65		151
-1 SEM	70		156
+1SEM	81		169
+2 SEMs	86		175

Cut Scores within 1 and 2 SEMs of the Recommended Cut Score – Multistate Study (Panel II)

Recommended Cut Score	77	Scale Score Equivalent	164
-2 SEMs	68		154
-1 SEM	73		160
+1SEM	82		170
+2 SEMs	87		176

As of November 9, 2009, ETS reported that seven states, the District of Columbia, and the Virgin Islands have set cut scores for the revised SLLA. Mississippi set their cut score at 169, and Kansas set their score at 165. Six states (Indiana, Missouri, New Jersey, North Carolina, and Utah) and the District of Columbia set the cut score at 163. The Virgin Islands’ cut score is 156.

The Advisory Board on Teacher Education and Licensure (ABTEL) reviewed the studies and cut scores established by other states (Appendix D) at its September 21, 2009, meeting. The members decided to delay making a recommendation for a cut score until the November 16, 2009, ABTEL meeting in order to review candidates’ scores from the first national administration of the *SLLA* (Appendix E). On November 16, the Advisory Board recommended a cut score of 161 for the *School Leaders Licensure Assessment* which is one Standard Error of Measurement above the Virginia panel’s recommended score. The recommendation was made with the caveat that the passing rates for the SLLA be reviewed after three test administrations of the test in Virginia. The Board of Education has the authority to set the final cut score for the revised *SLLA* assessment.

Superintendent's Recommendation:

The Superintendent of Public Instruction recommends that the Board of Education receive the

recommendation of the Advisory Board on Teacher Education and Licensure, approve a cut score for the *SLLA*, and review the passing rates after three test administrations of the *School Leaders Licensure Assessment* in Virginia.

Impact on Resources:

Costs associated with the administration of the *School Leaders Licensure Assessment (SLLA)* will be incurred by the Educational Testing Service. Prospective school leaders will be required to pay a fee for test administration and reporting results to the Virginia Department of Education. The cost for the revised *SLLA* assessment is \$375, including a nonrefundable \$50 registration fee.

Timetable for Further Review/Action:

The Department of Education will notify school divisions and institutions of higher education of the Board of Education's decision.

APPENDIX A

Standard Setting Report
School Leaders Licensure Assessment (SLLA)
March 24-25, 2009
Richmond, Virginia

Standard Setting Report

School Leaders Licensure Assessment (SLLA)

March 24 - 25, 2009
Richmond, Virginia

Conducted by
Educational Testing Service
Princeton, New Jersey

School Leaders Licensure Assessment (SLLA)

STANDARD SETTING

Introduction

A standard setting study was conducted on March 24-25 for the School Leaders Licensure Assessment (SLLA) used to award an Administrative Services Credential in Virginia. Educational Testing Service (ETS) conducted the standard setting study on behalf of the Virginia Department of Education (VDOE) for the revised version of the SLLA, which will be administered in Virginia for the first time in January 2010.

The purposes of the study were to (a) recommend the minimum SLLA score judged necessary to award an Administrative Services Credential and (b) confirm the importance of the SLLA content specifications for entry-level school (education) leaders in Virginia. The Office of Teacher Education and Licensure (in the VDOE) will submit the standard setting panel's recommended passing score, or cutscore, to the Advisory Board on Teacher Education and Licensure (ABTEL) for consideration. The ABTEL will forward a recommendation to the Virginia State Board of Education (VSBE); the VSBE sets the final, operational cutscore on the SLLA.

School Leaders Licensure Assessment

The purpose and structure of the SLLA are described in the School Leaders Licensure Assessment *Test at a Glance* (ETS, in press). In brief, the assessment measures whether entry-level school (education) leaders have the knowledge believed necessary for competent professional practice. The content of the assessment was defined by a National Advisory Committee of expert practitioners and preparation faculty and confirmed by a national survey of the field. The content is aligned with the Interstate School Leaders Licensure Consortium (ISLLC) Standards for School Leaders.

The four-hour SLLA is divided into two separately timed sections:

- **Section I** (2 hours 20 minutes) – 100 multiple-choice questions (80 operational and 20 pre-test).
- **Section II** (1 hour 40 minutes) – Seven constructed-response questions calling for written answers based on scenarios and sets of documents that an education leader might encounter. Candidates are required to analyze situations and data, to propose appropriate courses of action, and to provide rationales for their proposals.

Candidate scores on the two sections are weighted such that Section I contributes 70% of the overall SLLA score and Section II contributes 30%. The total number of raw points that may be earned on the

SLLA is 114 (80 points from the multiple-choice section and approximately 34 points from the constructed-response section). The reporting scale for the SLLA ranges from 100 to 200 scaled points.

Committee Members

A panel of 15 experts in school leadership participated in the standard-setting study. The Office of Teacher Education and Licensure recruited panelists to represent a range of professional perspectives, including principals, assistant principals and university faculty. Table 1 presents the demographic characteristics of the 15 panelists. In brief, eight panelists were principals, two were assistant principals, and five were college faculty. Nine panelists were White, five were African American, and one was Native American. Six panelists were female. All panelists reported being certified school leaders in Virginia. Twelve panelists had between 4 and 11 years of experience as a building-level administrator or central office supervisor. (See Appendix A for a list of the panelists.)

Process and Method

Prior to the Panel Meeting

The panelists were sent an e-mail explaining the purpose of the standard-setting study and the planned agenda (Appendix B), and requesting that they review the test content specifications for the SLLA (included in the *SLLA Test at a Glance*, which was attached to the e-mail). The purpose of the review was to familiarize the panelists with the general structure and content of the assessment.

During the Panel Meeting

The standard-setting study began with a welcome and introduction by Dr. James Lanham, Director of Licensure, of the VDOE's Office of Teacher Education and Licensure. Dr. Richard Tannenbaum, a director of research from ETS, then explained how the SLLA was developed, provided an overview of standard setting, and presented the agenda for the study. Dr. Tannenbaum served as the lead facilitator for the standard-setting meeting.

The first activity was for the panelists to "take the test." (Each panelist had signed a nondisclosure form.) The panelists were given approximately two hours to respond to the multiple-choice questions and to sketch responses to the constructed-response questions. The panelists had access to the answer key for the multiple-choice questions and access to both the general and question-specific rubrics for the constructed response questions. The purpose of "taking the test" was for the panelists to become familiar with the test format, content and its difficulty.

The panelists then engaged in a discussion of the major content areas being addressed by the SLLA; they were also asked to remark on any questions that they thought would be particularly challenging for entering school leaders, and questions that addressed content that would be particularly important for entering school leaders.

Following this discussion, the panelists defined the Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of skills believed necessary to be a qualified building-level administrator or central office supervisor in Virginia. The JQC definition is the operational definition of the cutscore. The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC. The panelists were split into three groups of five and each group was asked to write down its definition of a JQC. Each group referred to the Interstate School Leaders Licensure Consortium (ISLLC) 2008 standards and the SLLA *Test at a Glance* to guide their definition. Each group posted its definition on chart paper and a full-panel discussion occurred to reach consensus on a final definition (Appendix C).

The standard-setting process for the SLLA was conducted for the overall test, though one standard-setting approach was implemented for Section I (multiple-choice questions) and another approach was implemented for Section II (constructed-response questions). The recommended passing score for the SLLA is the weighted sum of the interim cutscores recommended by the panelists for each section. These approaches are described next, followed by the results of the standard-setting study.

Standard Setting for Section I (Multiple-Choice Questions). A probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) was used for Section I (multiple-choice questions). In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC would answer it correctly. Panelists made their judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1. The lower the value, the less likely it is that a JQC would answer the question correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly. The panelists were asked to approach the judgment process in two stages. First, they reviewed the definition of the JQC and the question and decided if, overall, the question was difficult for the JQC, easy for the JQC, or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rule of thumb to guide their decision:

- difficult questions for a JQC were in the 0 to .30 range;
- easy questions were in the .70 to 1 range; and
- moderately difficult/easy questions were in the .40 to .60 range.

The second decision was for panelists to decide how they wanted to refine their judgment within the range. For example, if a panelist thought that a question was easy for a JQC, the initial decision located the question in the .70 to 1 range. The second decision was for the panelist to decide if the likelihood of answering it correctly was .70, .80, .90, .95, or 1.0. The two-stage decision-process was implemented to reduce the cognitive load placed on the panelists. The panelists practiced making their standard-setting judgments on the first three questions.

The panelists engaged in three rounds of judgments. The Round 1 feedback provided to the panel included each panelist's (listed by ID number) recommended cutscore for the section and the panel's average recommended cutscore, highest and lowest cutscore, and standard deviation. Following discussion, the panelists' judgments were displayed for each question. The panelists' judgments were summarized by the three general difficulty levels (0 to .30, .40 to .60, and .70 to 1) and the panel's average question judgment was provided. Questions were highlighted to show when panelists converged in their judgments (10 of 15 located a question in the same difficulty range) or diverged in their judgments. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2). This process was repeated for the Round 2 judgments to inform their final (Round 3) judgments.

Standard Setting for Section II (Constructed-Response Questions). An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for Section II (constructed-response questions). In this approach, for each question, a panelist decides on the assigned score value that would most likely be earned by a JQC. The basic process that each panelist followed was first to review the definition of the JQC that was agreed upon and then to review the question and the rubrics (general and specific) for that question. The general rubric for a question defines holistically the quality of the evidence that would merit a response earning a 3 (*thorough understanding*), 2 (*basic/general understanding*), 1 (*limited understanding*), or 0 (*little or no understanding*). Each question-specific rubric provides examples of evidence that would support earning a 3, 2, 1, or 0. During this review, each panelist independently considered the level of knowledge and/or skill required to respond to the question and the features of a response that would earn 3, 2, 1, or 0 points, as defined by the rubrics.

A test taker's response to a constructed-response question is independently scored by two raters, and the sum of the raters' scores is the assigned score¹; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). Each panelist decided on the score most likely to be earned by a JQC from the following possible values: 0, 1, 2, 3, 4, 5, 6. A score of 3, for example, would mean that a panelist thought that a JQC would most likely earn between a 1 and a 2 from the two raters. A 5 would mean that a JQC would most likely earn between a 2 and a 3 from the two raters. For each of the seven constructed-response questions, panelists recorded the score (0 through 6) that a JQC would most likely earn.

The Section II (constructed-response questions) score is weighted to contribute 30% of the total test score; the Section I (multiple-choice question) score is weighted to contribute 70%. The facilitator explained the reason and process for weighting the contributions of the two sections to the panel. The

¹ If the two raters' scores differ by more than one point (non-adjacent), the Chief Reader for that question assigns the score, which is then doubled.

feedback to the panelists, described next, included both their judgments before and after the weightings so they could more easily see the relationship between their explicit standard-setting judgments (pre-weighting) and the post-weighting values used to compute the recommended cutscore.

Consistent with the standard-setting process used for Section I, the panelists engaged in three rounds of judgments for Section II. After each round, the judgments of each panelist were summarized and projected for the panel to see and discuss. Each panelist's recommended cutscore for Section II was displayed as was the panel's average recommended cutscore, highest and lowest cutscore, and standard deviation. The panelists' judgments also were displayed for each question. The panelists participated in a general discussion of the results. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments.

Judgment of SLLA Content Specifications

Following the three-round standard setting process, panelists judged the importance of the knowledge and/or skills stated or implied in the SLLA content specifications for the job of an entry-level school leader in Virginia. These judgments addressed the perceived content-based validity of the SLLA. Judgments were made using a four-point Likert scale — *very important*, *important*, *slightly important*, and *not important*. The panel first judged the importance of the first content category— *Vision and Goals* — and its three sub-categories. As a group, the panel discussed their judgments and were allowed to revise their judgments following the discussion. The panel independently judged the remaining categories and sub-categories.

Results

Initial Evaluation Forms

The panelists completed two initial evaluation forms, once after they were trained in how to make their standard-setting judgments for Section I (multiple-choice questions), and once after they were trained to make their judgments for Section II (constructed-response questions). The primary information collected from these forms was the panelists indicating if they had received adequate training to make their standard-setting judgments and were ready to proceed. All 15 panelists indicated that they were prepared to make their judgments.

Summary of Standard Setting Judgments by Round

Tables 2 through 4 present a summary of each round of standard-setting judgments for Section I (multiple-choice questions), Section II (constructed-response questions) and the overall test, respectively. The numbers in each table reflect the recommended cutscore—the number of raw points needed to “pass” the section or test—of each panelist for each of the three rounds. Note that the SLLA reports a single,

overall score and that the panel will be recommending a single cutscore for the weighted composite of Sections I and II. The separate “cutscores” for the two sections is an intermediate step in calculating the overall cutscore. For Section II, both the direct standard-setting judgments (“Raw Score”) and the weighted judgments (“Wt. Cutscore”) are presented. The panel’s average recommended cutscore and highest and lowest cutscores are reported, as are the standard deviation (SD) of panelists’ cutscores and the standard error of judgment (SEJ). The SEJ is one way of estimating the reliability of the judgments. It indicates how likely it would be for other panels of educators similar in make-up, experience, and standard-setting training to the current panel to reach the same cutscore. A comparable panel’s cutscore would be within 1 SEJ of the current average cutscore 68 percent of the time and within 2 SEJs 95 percent of the time.

Round 1 judgments are made without discussion among the panelists. The most variability in judgments, therefore, is typically present in the first round. Round 2 and Round 3 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease — indicating convergence among the panelists’ judgments — was observed for the SLLA.

The Round 3 average score for each section is summed to arrive at the SLLA recommended cutscore (passing score). It should be noted, however, that separate Section I and Section II scores are not reported for the SLLA; only an overall score is reported. Therefore, there is no required minimum section score that must be obtained in order to pass the SLLA. The total test cutscore is compensatory, in that as long as the total SLLA cutscore is met or exceeded, the candidate has passed. The panels’ cutscore recommendation for the SLLA is 67.24 (see Table 4); because this value is greater than 67, the next highest whole number, or 68, is considered the functional recommended cutscore. This value of 68 represents approximately 60% of the total available 114 raw points that could be earned on the SLLA. The scaled score associated with 68 raw points is 154².

Table 5 presents the standard error of measurement (SEM) around the recommended cutscore of 68 points. A standard error represents the uncertainty associated with a test score. The scaled scores associated with 1 and 2 SEMs are provided. The standard error provided is an estimate, given that the SLLA has not yet been administered.

Summary of Specification Judgments

Panelists judged the extent to which the knowledge and/or skills reflected by the SLLA content specifications were important for entry-level school leaders in Virginia. Panelists rated the six content categories and their accompanying sub-categories, on a four-point scale ranging from *very important* to *not important*. The panelists’ ratings are summarized in Table 6. Overall, the majority of panelists (93%

² For reference purposes, if the recommended raw cutscore were 67 points, the scaled score would be 153.

or greater) judged all six content categories to be either *very important* or *important* for entry-level school leaders in Virginia. In particular, panelists judged *Vision and Goals*, *Teaching and Learning*, and *Ethics and Integrity* to be *very important*. A majority of panelists (80% or greater) also judged each of the sub-categories to be *very important* or *important*. Three panelists, nonetheless, rated the *Maximizing Community Resources* and *Professional Influence* sub-categories as *slightly important*.

Final Evaluations

The panelists completed an evaluation form at the conclusion of the standard-setting study. The evaluation form asked the members to provide feedback about the quality of the standard-setting implementation, the factors that influenced their decisions, and their comfort with (acceptance of) the SLLA recommended cutscore. Table 7 presents the results of the final evaluations.

All panelists confirmed that they understood the purpose of the study; that the facilitator's instructions and explanations were clear; that they were prepared to make their standard setting judgments; and that the standard-setting process was easy to follow. The panelists reported that their own professional experience and the definition of the JQC most influenced their standard-setting judgments. The majority of panelists (11 out of 15 or 73%) reported being *very comfortable* with the SLLA recommended cutscore, and all panelists reported that the cutscore was *about right*.

Summary

The purposes of this standard setting study were to (a) recommend the minimum SLLA score judged necessary to award an Administrative Services Credential and (b) confirm the importance of the SLLA content specifications for entry-level school leaders in Virginia. A panel of 15 principals, assistant principals and college faculty was assembled to make the cutscore recommendation and to review the SLLA content specifications. Standard setting was conducted using a probability-based Angoff approach (for the multiple-choice section) and an Extended Angoff approach (for the constructed-response section). Section-level minimum scores were constructed and a weighted sum was computed. The average across panelists was 67.24; because this value is greater than 67, the next highest whole number, or 68, is considered the functional recommended cutscore. This value of 68 represents approximately 60% of the total available 114 raw points that could be earned on the SLLA. The scaled score associated with 68 raw points is 154. The panel confirmed that the knowledge and/or skills stated or implied in the SLLA content specifications were important for entry-level school leaders in Virginia. The results of the evaluation surveys (initial and final) support the quality of the standard-setting implementation.

References

- Brandon, P.R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education, 17*, 59-88.
- Cizek, G. J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Educational Testing Service. (in press). *School Leaders Licensure Assessment: Test at a glance*. Princeton, NJ: Author.
- Hambleton, R. K., & Pitoniak, M.J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 433-470). Westport, CT: American Council on Education/Praeger.
- Hambleton, R. K., & Plake, B.S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41-55.

Table 1. Committee Member Demographics

	N	Percent
Group you are representing		
Principal	8	53%
Assistant Principal	2	13%
College Faculty	5	33%
Other	0	0%
Race		
African American or Black	5	33%
Alaskan Native or American Indian	1	7%
Asian or Asian American	0	0%
Native Hawaiian or Other Pacific Islander	0	0%
White	9	60%
Hispanic	0	0%
Gender		
Female	6	40%
Male	9	60%
Are you certified as a school leader in Virginia?		
No	0	0%
Yes	15	100%
Are you currently a school leader in Virginia?		
No	4	27%
Yes	11	73%
Are you currently mentoring another school leader?		
No	6	40%
Yes	9	60%
How many years of experience do you have as a school leader in Virginia?		
3 years or less	0	0%
4 - 7 years	8	53%
8 - 11 years	4	27%
12 - 15 years	0	0%
16 years or more	3	20%
For which education level are you currently a school leader?		
Pre K - Kindergarten	0	0%
K - Grade 5	2	13%
Grades 6 - 8	3	20%
Grades 9 - 12	4	27%
Grades K-12	1	7%
Higher Education	5	33%
School Setting		
Urban	6	40%
Suburban	7	47%
Rural	2	13%

Table 2. Section I (Multiple-Choice Questions): Summary of Each Round of Judgments

Panelist	Round 1 Cutscore	Round 2 Cutscore	Round 3 Cutscore
1	55.75	55.75	55.75
2	53.45	52.25	52.25
3	40.40	40.40	40.70
4	51.20	49.45	49.55
5	54.60	52.70	52.30
6	45.55	45.05	44.85
7	56.55	55.30	55.10
8	52.00	50.90	50.50
9	48.75	47.25	46.75
10	36.50	39.10	39.30
11	34.40	35.00	37.90
12	48.30	47.50	47.20
12	51.40	51.80	52.70
14	47.80	45.30	45.20
15	38.95	35.45	35.85
Average	47.71	46.88	47.06
SD	7.10	6.76	6.34
SEJ	1.83	1.75	1.64
Highest	56.55	55.75	55.75
Lowest	34.40	35.00	35.85

Table 3. Section II (Constructed-Response Questions): Summary of Each Round of Judgments

Member	Round 1		Round 2		Round 3	
	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore
1	32.00	26.11	32.00	26.11	32.00	26.11
2	28.00	22.85	23.00	18.77	23.00	18.77
3	19.00	15.50	19.00	15.50	19.00	15.50
4	24.00	19.58	24.00	19.58	24.00	19.58
5	26.00	21.22	26.00	21.22	26.00	21.22
6	24.00	19.58	23.00	18.77	23.00	18.77
7	30.00	24.48	28.00	22.85	28.00	22.85
8	33.00	26.93	29.00	23.66	28.00	22.85
9	23.00	18.77	24.00	19.58	24.00	19.58
10	26.00	21.22	26.00	21.22	26.00	21.22
11	30.00	24.48	30.00	24.48	30.00	24.48
12	14.00	11.42	19.00	15.50	19.00	15.50
13	25.00	20.40	26.00	21.22	26.00	21.22
14	27.00	22.03	25.00	20.40	25.00	20.40
15	19.00	15.50	18.00	14.69	18.00	14.69
Average	25.33	20.67	24.80	20.24	24.73	20.18
SD	5.18	4.23	4.07	3.32	4.01	3.27
SEJ	1.34	1.09	1.05	0.86	1.03	0.84
Highest	33.00	26.93	32.00	26.11	32.00	26.11
Lowest	14.00	11.42	18.00	14.69	18.00	14.69

Note: The maximum raw score for Section II is 42 points, with a maximum weighted score of 34.27.

Table 4. Recommended SLLA Cutscores: Summary of Each Round of Judgments

Panelist	Round 1	Round 2	Round 3
	Overall Cutscore	Overall Cutscore	Overall Cutscore
1	81.86	81.86	81.86
2	76.30	71.02	71.02
3	55.90	55.90	56.20
4	70.78	69.03	69.13
5	75.82	73.92	73.52
6	65.13	63.82	63.62
7	81.03	78.15	77.95
8	78.93	74.56	73.35
9	67.52	66.83	66.33
10	57.72	60.32	60.52
11	58.88	59.48	62.38
12	59.72	63.00	62.70
12	71.80	73.02	73.92
14	69.83	65.70	65.60
15	54.45	50.14	50.54
Average	68.38	67.12	67.24
SD	9.37	8.63	8.34
SEJ	2.42	2.23	2.15
Highest	81.86	81.86	81.86
Lowest	54.45	50.14	50.54

Table 5. Cutscores within 1 and 2 SEMs of the Recommended Cutscore

Recommended Cutscore (SEM)	Scale Score Equivalent
68 (5.12)	154
- 2 SEMs	143
-1 SEM	149
+1 SEM	161
+ 2 SEMs	167

Note: Consistent with the recommended cutscore, the cutscores at the different SEMs have been rounded to the next highest whole number.

Table 6. SLLA Specification Judgments

	Very Important		Important		Slightly Important		Not Important	
	N	%	N	%	N	%	N	%
I. Vision and Goals (20%)	12	80%	2	13%	1	7%	0	0%
A. Vision and goals for teaching and learning	12	80%	2	13%	1	7%	0	0%
B. Shared commitments to implement the vision and goals	9	60%	5	33%	1	7%	0	0%
C. Continuous improvement toward the vision and goals	11	73%	3	20%	1	7%	0	0%
II. Teaching and Learning (30%)	15	100%	0	0%	0	0%	0	0%
A. Building a professional culture	13	87%	2	13%	0	0%	0	0%
B. Rigorous curriculum and instruction	13	87%	2	13%	0	0%	0	0%
C. Assessment and accountability	13	87%	2	13%	0	0%	0	0%
III. Managing Organizational Systems and Safety (10%)	8	53%	6	40%	1	7%	0	0%
A. Managing Operational Systems	4	27%	11	73%	0	0%	0	0%
B. Aligning and obtaining fiscal and human resources	6	40%	8	53%	1	7%	0	0%
C. Protecting the welfare and safety of students and staff	14	93%	1	7%	0	0%	0	0%
IV. Collaborating with Key Stakeholders (15%)	8	53%	7	47%	0	0%	0	0%
A. Collaborate with families and other community members	9	60%	6	40%	0	0%	0	0%
B. Community interests and needs	7	47%	8	53%	0	0%	0	0%
C. Maximizing community resources	3	20%	9	60%	3	20%	0	0%
V. Ethics and Integrity (15%)	14	93%	1	7%	0	0%	0	0%
A. Ethical and legal behavior	14	93%	1	7%	0	0%	0	0%
B. Personal values and beliefs	13	87%	1	7%	1	7%	0	0%
C. High standards for self and others	14	93%	1	7%	0	0%	0	0%
VI. The Education System (10%)	6	40%	8	53%	1	7%	0	0%
A. Professional influence	4	27%	8	53%	3	20%	0	0%
B. Managing local decisions within the larger educational policy environment	4	27%	9	60%	2	13%	0	0%

Table 7. Final Evaluation

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	%	N	%	N	%	N	%
I understood the purpose of this study.	12	80%	3	20%	0	0%	0	0%
The instructions and explanations provided by the facilitator were clear.	14	93%	1	7%	0	0%	0	0%
The training in the standard setting methods was adequate to give me the information I needed to complete my assignment.	13	87%	2	13%	0	0%	0	0%
The explanation of how the recommended cut scores are computed was clear.	13	87%	2	13%	0	0%	0	0%
The opportunity for feedback and discussion between rounds was helpful.	12	80%	3	20%	0	0%	0	0%
The process of making the standard setting judgments was easy to follow.	12	80%	3	20%	0	0%	0	0%

	Very Influential		Somewhat Influential		Not Influential	
	N	%	N	%	N	%
How influential was each of the following factors in guiding your standard setting judgments?						
The definition of the Just Qualified Candidate	10	67%	5	33%	0	0%
The between-round discussions	8	53%	6	40%	1	7%
The cut scores of other panel members	2	13%	12	80%	1	7%
My own professional experience	12	80%	3	20%	0	0%

	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	%	N	%	N	%	N	%
Overall, how comfortable are you with the panel's recommended cut score?	11	73%	4	27%	0	0%	0	0%

	Too Low		About Right		Too High	
	N	%	N	%	N	%
Overall, the panel's recommended cut score for the SLLA test is:	0	0%	15	100%	0	0%

Appendix A

List of Panelists

Daryl Chesley	James Monroe High School
Karen S. Crum	Old Dominion University
Vincent M. Darby Sr.	Norfolk Public Schools
Vicki P. Duling	Franklin Sherman Elementary School, Fairfax County Public Schools
Beverly D. Epps	University of Mary Washington
William F. Floro	Radford University
Anthony Francis	Jefferson Forest High School
Glenn L. Koonce	Regent University
Janice Koslowski	Loudoun County Public Schools
Mark Mear	Lynchburg City Schools
Earl F. Newby	Virginia State University
A. Katrise Perera	Henrico County Public Schools
Joey H. Phillips	Virginia Beach City Public Schools
Tiffany Demarest Sanzo	Kecoughtan High School, Hampton City Schools
Lawrence P. Whiting	Brunswick County Public Schools

Appendix B

Agenda

March 24, 2009

- 8:00 – 8:15 AM Welcome and Introductions
- 8:15 – 8:45 Overview of Standard Setting and Workshop Events
Sign nondisclosure and complete biographical information form
- 8:45 – 10:45 Review of the SLLA
Break as needed
- 10:45 – 11:05 Discuss SLLA
What is being measured?
What does an entering school (education) leader need to know and do?
- 11:05 – 12:15 PM Define Knowledge/Skills of Just Qualified Candidate
- 12:15 – 1:00 Lunch
- 1:00 – 1:45 Standard Setting Training for Multiple-choice items
Practice judgments -- first 3 items
Complete training evaluation form
- 1:45 – 3:15 Round 1 Standard Setting Judgments for Multiple-choice Items
Break as needed
- 3:15 – 3:45 Standard Setting Training for Constructed-response items
Practice judgments – first item
Complete training evaluation form
- 3:45 – 4:45 Round 1 Standard Setting Judgments for Constructed-response items
- 4:45 – 5:00 Collect Materials; End of Day 1

Agenda

March 25, 2009

- 8:00 – 8:15 AM Questions From Day 1? Overview of Day 2
- 8:15 – 10:15 Round 1 Feedback and Discussion
- 10:15 – 10:45 Round 2 Standard Setting Judgments MC and CR Items
- 10:45 – 11:00 Data Entry; Break
- 11:00 – 12:00 PM Round 2 Feedback and Discussion
- 12:00 – 12:45 Lunch
- 12:45 – 1:00 Round 3 Standard Setting Judgments MC and CR Items
- 1:00 – 1:30 Specification Judgment Training
- Practice judgments – first specification: *Vision and Goals*
- 1:30 – 2:00 Complete Specification Judgments
- 2:00 – 2:15 Feedback on Round 3 Recommended Cut Score
- 2:15 – 2:30 Complete Final Evaluation
- 2:30 – 2:45 Collect Materials; End of Study

Appendix C

Definition of a Just Qualified Candidate

- Acts with integrity, fairness, and in an ethical manner.
- Analyzes and synthesizes data for instructional purposes and school processes to make decisions.
- Has the ability to evaluate and support instructional methods.
- Has basic knowledge of fiscal management.
- Has basic knowledge of human resource management.
- Has the ability to lead a diverse student population – meet the needs of all students.
- Knows how to evaluate programs for success.
- Knows how to communicate and collaborate effectively with all stakeholders.
- Knows how to deal with conflict.
- Knows how to use technology for instructional improvement.
- Understands that the school vision is a process.
- Knows how to create and maintain a safe school environment.
- Understands applicable local, state, and federal laws and guidelines as they affect student learning

APPENDIX B

Multi-State Standard Setting Report School Leaders Licensure Assessment (SLLA)

*April 21-22, 2009
Baltimore, Maryland*

*May 12-13, 2009
St. Louis, Missouri*



Multi-State Standard Setting Report School Leaders Licensure Assessment (SLLA)

April 21 - 22, 2009

Baltimore, Maryland

May 12 - 13, 2009

St. Louis, Missouri

Conducted by
Educational Testing Service
Princeton, New Jersey

Introduction

Research staff from Educational Testing Service (ETS) designed and conducted two multi-state standard setting studies for the revised School Leaders Licensure Assessment (SLLA). The primary purpose of the studies was to provide state departments of education with recommendations regarding a passing score or cutscore. Each study also collected content-related validity evidence of the importance of the SLLA content specifications for entry-level school (education) leaders. Two non-overlapping panels totaling 46 practicing school leaders and college faculty who prepare school leaders participated. The participants represented 17 states (see Table 1) that currently use the SLLA or plan to use the assessment; participants were selected by their respective state departments of education.

Table 1. Participating States (and number of panelists)

• Arkansas (2)	• Mississippi (4)
• California (2)	• Missouri (4)
• Connecticut (4)	• New Jersey (4)
• Indiana (1)	• North Carolina (2)
• Kansas (1)	• Tennessee (2)
• Kentucky (3)	• Utah (3)
• Louisiana (4)	• Washington, DC (2)
• Maine (2)	• Wyoming (2)
• Maryland (4)	

NOTE: Indiana and Kansas were represented on Panel 1 and North Carolina was represented on Panel 2. All other listed states were represented on both panels.

The use of two multi-state panels (a) enabled each state to be represented in the passing score recommendation and (b) provided an opportunity to replicate the standard setting process, which reinforces the quality of the passing score recommendation. The training provided to panelists and study materials were consistent across the two panels with the exception of defining the “just qualified candidate (JQC).” The JQC is the borderline test taker, or the candidate with the minimum level of knowledge and/or skills believed necessary to pass the assessment. This is the operational definition of the passing score. The standard setting process is designed to identify the assessment score that aligns with this definition. To assure that both panels were using the same frame of reference when making their standard setting judgments, the JQC defined by the first panel served as the definition for the second panel. The second panel completed a thorough review of the definition to allow panelists to discuss and internalize the definition. The processes for developing the definition (with Panel 1) and reviewing/internalizing the definition (with Panel 2) are described later and the Just Qualified Candidate definition is presented in Appendix A.

The panels were convened on April 21-22, 2009 in Baltimore, Maryland and on May 12-13, 2009 in St. Louis, Missouri. The results for each panel and results combined across panels are summarized in the following report. This report will be provided to each of the represented state departments of education. In each state, the department of education, the state board of education or a designated educator licensure board is responsible for establishing the final SLLA passing score in accord with applicable state regulations.

The first national administration of the revised School Leaders Licensure Assessment will be in September 2009. The current version of the SLLA will be phased out, with the last national administration in June 2009.

School Leaders Licensure Assessment

The School Leaders Licensure Assessment *Test at a Glance* (ETS, in press) describes the purpose and structure of the SLLA. In brief, the assessment measures whether entry-level school (education) leaders have the knowledge and/or skills believed necessary for competent professional practice. A National Advisory Committee of expert practitioners and preparation faculty defined the content of the SLLA and a national survey of the field confirmed the content. The content is aligned with the Interstate School Leaders Licensure Consortium (ISLLC) Standards for School Leaders.

The four-hour SLLA is divided into two separately timed sections:

- **Section I** (2 hours 20 minutes) – 100 scenario-based multiple-choice questions (80 operational and 20 pre-test).
- **Section II** (1 hour 40 minutes) – Seven constructed-response questions calling for written answers based on scenarios and sets of documents that a school leader might encounter.

Candidate scores on the two sections are weighted such that Section I contributes 70% of the overall SLLA score and Section II contributes 30%. The total number of raw points that may be earned on the SLLA is 114 (80 points from the multiple-choice section and approximately 34 points from the constructed-response section). The reporting scale for the SLLA ranges from 100 to 200 scaled-score points.

Expert Panels

Panel 1 included 23 school leaders and college faculty who prepare school leaders representing 16 states. The various state departments of education recruited panelists to represent a range of professional perspectives. In brief, ten panelists were principals, two were assistant principals, and eight were college faculty. Fourteen panelists were White, seven were African American, one was Hispanic

and one was Asian American. Eleven panelists were female. Twenty-one panelists reported being certified school leaders in their states. Nine panelists had between 4 and 11 years of experience as a school leader.

Panel 2 also included 23 school leaders and college faculty who prepare school leaders representing 15 states. As with Panel 1, the various state departments of education recruited panelists to represent a range of professional perspectives. Twelve panelists were principals, four were assistant principals, and five were college faculty. Eleven panelists were White, eight were African American, two were Hispanic and one each was Asian American and Alaskan Native/American Indian. Thirteen panelists were female. Twenty-one panelists reported being certified school leaders in their states. Fourteen panelists had between 4 and 11 years of experience as a school leader

Table 1 presents the demographic characteristics of the two panels. Lists of panel members and their affiliations are presented in Appendix B.

Process and Method

The training provided to panelists and study materials were consistent across panels. Any differences between panels (e.g., defining the Just Qualified Candidate) are highlighted.

Prior to each study, the panelists were sent an e-mail explaining the purpose of the standard setting study and the planned agenda (Appendix C), and requesting that they review the test content specifications for the SLLA (included in the *SLLA Test at a Glance*, which was attached to the e-mail). The purpose of the review was to familiarize the panelists with the general structure and content of the assessment.

Each standard setting study began with a welcome and introduction by Drs. Richard Tannenbaum and Clyde Reese, ETS researchers in the Center for Validity Research. Dr. Tannenbaum, lead facilitator for the study, then explained how the SLLA was developed, provided an overview of standard setting, and presented the agenda for the study.

Reviewing the SLLA

The first activity was for the panelists to “take the test.” (Each panelist had signed a nondisclosure form.) The panelists were given approximately two hours to respond to the multiple-choice questions and to sketch responses to the constructed-response questions. The panelists had access to the answer key for the multiple-choice questions and access to both the general and question-specific rubrics for the constructed-response questions. The purpose of “taking the test” was for the panelists to become familiar with the test format, content and its difficulty.

The panelists then engaged in a discussion of the major content areas being addressed by the SLLA. They were also asked to remark on any questions that they thought would be particularly challenging for entering school leaders, and questions that addressed content that would be particularly important for entering school leaders.

Defining the JQC

Following the review of the SLLA, panelists defined the knowledge and/or skills expected of a Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of knowledge and/or skills believed necessary to be a qualified school leader. The JQC definition is the operational definition of the passing score (cutscore). The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC.

In Panel 1, the panelists were split into four groups and each group was asked to write down its definition of a JQC. Each group referred to the Interstate School Leaders Licensure Consortium (ISLLC) 2008 standards, the SLLA *Test at a Glance*, and an example JQC description for the SLLA (developed by one state) to guide their definition. Each group posted its definition on chart paper and a full-panel discussion occurred to reach consensus on a final definition.

In Panel 2, the panelists began with the definition of the JQC developed by the first panel. Given that the SLLA multi-state standard setting was designed to replicate processes and procedures across the two panels, it was important that both panels use the same JQC definition to frame their judgments¹. For Panel 2, the panelists reviewed the (Panel 1) JQC definition and any ambiguities were discussed and clarified. The panelists then were split into four groups and each group developed three to four examples of behaviors or decisions they would expect of a JQC based on the definition. The examples were shared across groups and discussed. The purpose of the exercises was to have the panelists internalize the definition.

Minor edits were made to the existing JQC definition based on the discussions. For example, Knowledge Statement 9 from Panel 1 was “[the JQC] models principles of self-awareness, reflective practice, transparency, and loyalty to the shared vision.” Panel 2 deleted the phrase “loyalty to the shared vision,” believing that the notion of loyalty to the vision was subsumed by other statements and detracted from the other aspects of the knowledge statement. Panel 2 also modified Knowledge Statement 13 from the original, “[the JQC] knows and uses the basic tenets of teaching and learning (including the use of

¹Discussions regarding the “use of technology” occurred in Panel 1 during (a) the initial defining of the JQC and (b) Round 1 judgments. To reflect these discussions, language was added to the definition of the JQC that was carried forward to Panel 2. The parenthetical was added to the bullet “*Knows and uses the basic tenets of teaching and learning (including the use of technology to support teaching and learning).*”

technology to support teaching and learning),” to “knows and applies the basic tenets of teaching and learning (including the use of technology to support teaching and learning) as it impacts student achievement.” These changes helped to clarify the JQC definition for Panel 2, but did not alter the overall level of knowledge expected of the JQC. The JQC definition, with Panel 2 revisions highlighted, is presented in Appendix A.

Panelists’ Judgments

The standard-setting process for the SLLA was conducted for the overall assessment, though one standard-setting approach was implemented for Section I (multiple-choice questions) and another approach was implemented for Section II (constructed-response questions). Each panel’s passing score for the SLLA is the weighted sum of the minimum scores recommended by the panelists for each section. These approaches are described next, followed by the results from each standard-setting panel. The average of the two passing score recommendations is also provided.

Standard Setting for Section I (Multiple-Choice Questions). A probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) was used for Section I (multiple-choice questions). In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC would answer it correctly. Panelists made their judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1. The lower the value, the less likely it is that a JQC would answer the question correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly.

For each panel, the panelists were asked to approach the judgment process in two stages. First, they reviewed the definition of the JQC and the assessment question and decided if, overall, the question was difficult for the JQC, easy for the JQC, or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- difficult questions for a JQC were in the 0 to .30 range;
- easy questions were in the .70 to 1 range; and
- moderately difficult/easy questions were in the .40 to .60 range.

The second decision was for panelists to decide how they wanted to refine their judgment within the range. For example, if a panelist thought that a question was easy for a JQC, the initial decision located the question in the .70 to 1 range. The second decision was for the panelist to decide if the likelihood of answering it correctly was .70, .80, .90, .95, or 1.0. The two-stage decision-process was implemented to

reduce the cognitive load placed on the panelists. The panelists practiced making their standard-setting judgments on the first three multiple-choice questions.

The panelists engaged in three rounds of judgments. The Round 1 feedback provided to the panel included each panelist's (listed by ID number) recommended cutscore for the section and the panel's average recommended cutscore, highest and lowest cutscore, and standard deviation. Following discussion, the panelists' judgments were displayed for each question. The panelists' judgments were summarized by the three general difficulty levels (0 to .30, .40 to .60, and .70 to 1) and the panel's average question judgment was provided. Questions were highlighted to show when panelists converged in their judgments (approximately two-thirds of the panelists located a question in the same difficulty range) or diverged in their judgments. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2). This process was repeated for the Round 2 judgments to inform their final (Round 3) judgments.

Other than the definition of the JQC, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

Standard Setting for Section II (Constructed-Response Questions). An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for Section II (constructed-response questions). In this approach, for each question, a panelist decides on the assigned score value that would most likely be earned by a JQC. The basic process that each panelist followed was first to review the definition of the JQC and then to review the question and the rubrics (general and specific) for that question. The general rubric for a question defines holistically the quality of the evidence that would merit a response earning a 3 (*thorough understanding*), 2 (*basic/general understanding*), 1 (*limited understanding*), or 0 (*little or no understanding*). Each question-specific rubric provides examples of evidence that would support earning a 3, 2, 1, or 0. During this review, each panelist independently considered the level of knowledge and/or skill required to respond to the question and the features of a response that would earn 3, 2, 1, or 0 points, as defined by the rubrics.

A test taker's response to a constructed-response question is independently scored by two raters, and the sum of the raters' scores is the assigned score²; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). Each panelist decided on the

² If the two raters' scores differ by more than one point (non-adjacent), the Chief Reader for that question assigns the score, which is then doubled.

score most likely to be earned by a JQC from the following possible values: 0, 1, 2, 3, 4, 5, 6. A score of 3, for example, would mean that a panelist thought that a JQC would most likely earn between a 1 and a 2 from the two raters. A score of 5 would mean that a JQC would most likely earn between a 2 and a 3 from the two raters. For each of the seven constructed-response questions, panelists recorded the score (0 through 6) that a JQC would most likely earn. The panelists practiced making their standard-setting judgments on the first constructed-response question.

The Section II (constructed-response questions) score is weighted to contribute 30% of the total test score; the Section I (multiple-choice question) score is weighted to contribute 70%. The facilitator explained the reason and process for weighting the contributions of the two sections to the panel. The feedback to the panelists, described next, included both their judgments before and after the weightings so they could more easily see the relationship between their explicit standard-setting judgments (pre-weighting) and the post-weighting values used to compute the recommended cutscore.

Consistent with the standard-setting process used for Section I, the panelists engaged in three rounds of judgments for Section II. After each round, the judgments of each panelist were summarized and displayed for the panel to see and discuss. Each panelist's recommended cutscore for Section II was presented as was the panel's average recommended cutscore, highest and lowest cutscore, and standard deviation. The panelists' judgments also were displayed for each question. The panelists participated in a general discussion of the results. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2). This process was repeated for the Round 2 judgments to inform their final (Round 3) judgments.

As with Section I, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

Judgment of SLLA Content Specifications

Following the three-round standard setting process, each panel judged the importance of the knowledge and/or skills stated or implied in the SLLA content specifications for the job of an entry-level school leader. These judgments addressed the perceived content-based validity of the SLLA. Judgments were made using a four-point Likert scale — *very important*, *important*, *slightly important*, and *not important*. The panel first judged the importance of the first content category— *Vision and Goals* — and its three sub-categories. As a group, the panel discussed their judgments and were allowed to revise their

judgments following the discussion. Each panel independently judged the remaining categories and sub-categories.

Results

Initial Evaluation Forms

The panelists completed two initial evaluation forms, once after they were trained in how to make their standard-setting judgments for Section I (multiple-choice questions), and once after they were trained to make their judgments for Section II (constructed-response questions). The primary information collected from these forms was the panelists indicating if they had received adequate training to make their standard-setting judgments and were ready to proceed. On each panel, all panelists indicated that they were prepared to make their judgments.

Summary of Standard Setting Judgments by Round

Tables 2 through 4 present a summary of each round of standard-setting judgments for Section I (multiple-choice questions), Section II (constructed-response questions) and the overall assessment, respectively. Tables 2A, 3A, and 4A summarize results for Panel 1 and Tables 2B, 3B, and 4B summarize results for Panel 2. The numbers in each table reflect the recommended cutscores—the number of raw points needed to “pass” the section or assessment—of each panelist for each of the three rounds. Note that the SLLA reports a single, overall score and that each panel recommends a single cutscore for the weighted composite of Sections I and II. The separate “cutscores” for the two sections are intermediate steps in calculating the overall cutscore. For Section II, both the direct standard-setting judgments (“Raw Score”) and the weighted judgments (“Wt. Cutscore”) are presented. Each panel’s average recommended cutscore and highest and lowest cutscores are reported, as are the standard deviations (SD) of panelists’ cutscores and the standard errors of judgment (SEJ). The SEJ is one way of estimating the reliability of the judgments. It indicates how likely it would be for other panels of educators similar in make-up, experience, and standard-setting training to the current panels to reach the same cutscore. A comparable panel’s cutscore would be within 1 SEJ of the current average cutscore 68 percent of the time and within 2 SEJs 95 percent of the time.

Round 1 judgments are made without discussion among the panelists. The most variability in judgments, therefore, is typically present in the first round. Round 2 and Round 3 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease — indicating convergence among the panelists’ judgments — was observed for the SLLA in both panels.

The Round 3 average score for each section is summed to arrive at each panel's SLLA recommended cutscore (passing score). Although the passing score is based on the two sections, there is only an overall passing score requirement for the assessment. There are no required minimum section scores that must be obtained in order to pass the SLLA. The total assessment cutscore is compensatory, in that as long as the total SLLA cutscore is met or exceeded, the candidate has passed.

The panels' cutscore recommendation for the SLLA is 74.41 for Panel 1 and 76.16 for Panel 2 (see Tables 4A and 4B). The values were rounded to the next highest whole number to determine the functional recommended cutscores — 75 for Panel 1 and 77 for Panel 2. The values of 75 and 77 represent approximately 66% and 68%, respectively, of the total available 114 raw points that could be earned on the SLLA. The scaled scores associated with 75 and 77 raw points are 162 and 164, respectively.³

Tables 5A and 5B present the standard errors of measurement (SEM) around the recommended cutscores from each respective panel. A standard error represents the uncertainty associated with an assessment score. The scaled scores associated with 1 and 2 SEMs are provided. The standard errors provided are an estimate, given that the SLLA has not yet been administered. The SEM may be used to construct confidence intervals around a recommended cutscore. The confidence intervals around the Panel 1 and Panel 2 cutscore recommendations overlap, which means that the recommendations are not significantly different from one another. The two panels converged in their passing score recommendations, supporting that the standard setting outcome was successfully replicated.

Given the similarity of the passing score recommendations for the two panels, it is reasonable for state departments of education to consider the average of the two recommendations when making their operational passing score decision. The average of the two panel's recommendations is 75.29. The value was rounded to 76 (next highest raw score) to determine the functional recommended cutscore. The value of 76 represent approximately 67% of the total available 114 raw points that could be earned on the SLLA. The scaled score associated with 76 raw points is 163⁴. The estimated standard error associated with 76 raw points is 5.02. Raw score values (rounded to the next highest whole number) and [scaled score values] within 2 SEMs range from 66 [152] to 87 [176].

³ For reference purposes, if the recommended raw cutscore were 74 or 76 points, the scaled score would be 161 or 163, respectively.

⁴ For reference purposes, if the recommended raw cutscore was 75 points, the scaled score would be 162.

Summary of Specification Judgments

Panelists judged the extent to which the knowledge and/or skills reflected by the SLLA content specifications were important for entry-level school leaders. Panelists rated the six content categories and their accompanying sub-categories, on a four-point scale ranging from *very important* to *not important*. The panelists' ratings are summarized in Tables 6A (Panel 1) and 6B (Panel 2). For both panels, all panelists judged three of the six categories (Vision and Goals; Teaching and Learning; and Ethics and Integrity) to be either *very important* or *important* for entry-level school leaders; all but one panelist on one or both panels judged Managing Organization Systems and Safety; Collaborating with Key Stakeholders; and The Education System to be *very important* or *important*. For both panels, the categories Teaching and Learning, and Ethics and Integrity received the highest percentage of *very important* ratings. The majority of panelists on each panel (87% or greater; 20 of 23) also confirmed the importance of each of the sub-categories. The results across both panels support the content-based validity of the SLLA test content specifications.

Summary of Final Evaluations

The panelists on each panel completed an evaluation form at the conclusion of their standard setting study. The evaluation form asked the members to provide feedback about the quality of the standard-setting implementation, the factors that influenced their decisions, and their comfort with (acceptance of) the SLLA recommended cutscore. Tables 7A (Panel 1) and 7B (Panel 2) present the results of the final evaluations.

All panelists on each panel confirmed that they understood the purpose of the study; that the facilitators' instructions and explanations were clear; and that they were prepared to make their standard setting judgments. All but one panelists (on Panel 2) indicated that the standard-setting process was easy to follow. The panelists reported that their own professional experience and the definition of the JQC most influenced their standard-setting judgments.

There were some minor differences between the two panels when asked to respond to their level of comfort with their panel's recommended passing score and with their judgments regarding whether the recommended passing score was *too low*, *about right*, or *too high*. Nine panelists from Panel 1 reported being *very comfortable* with their panel's recommended passing score, and nine reported being *somewhat comfortable*. Thirteen panelists thought that the recommended passing score was *about right*, and nine thought it was *too low*. Fifteen panelists from Panel 2 reported being *very comfortable* with their panel's recommended passing score. Twenty panelists thought that the recommended passing score was *about right*, and three thought it was *too high*. These results suggest that taking the average of the two

recommended cutscores has perceived merit, as the average value of 76 is slightly higher than the recommended value from Panel 1 (75) and slightly lower than the recommended value of Panel 2 (77).

Summary

Two multi-state standard setting studies were conducted to recommend a passing score (cutoff) on the revised SLLA. Each study also collected content-related validity evidence of the importance of the SLLA content specifications for entry-level school (education) leaders. A total of 46 experts (e.g., school leaders and college faculty who prepare school leaders) representing 17 states participated.

Standard setting was conducted using a probability-based Angoff approach (for the multiple-choice section) and an Extended Angoff approach (for the constructed-response section). Section-level minimum scores were constructed and a weighted sum was computed. The cutoff recommendations for the SLLA were 74.41 for Panel 1 and 76.16 for Panel 2. The values were rounded to the next highest whole number to determine the functional recommended cutscores — 75 for Panel 1 and 77 for Panel 2. The values of 75 and 77 represent approximately 66% and 68%, respectively, of the total available 114 raw points that could be earned on the SLLA. The scaled scores associated with 75 and 77 raw points are 162 and 164, respectively.

Given the similarity of the passing score recommendations for the two panels, it is reasonable for state departments of education to consider the average of the two recommendations—75.29 or 76 rounded to the next highest whole number—when making their operational passing score decision. The scaled score associated with 76 is 163.

Both panels confirmed that the knowledge and/or skills stated or implied in the SLLA content specifications were important for entry-level school leaders. The results of the evaluation surveys (initial and final) from each panels support the quality of the standard-setting implementations.

References

- Brandon, P.R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education, 17*, 59-88.
- Cizek, G. J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Educational Testing Service. (in press). *School Leaders Licensure Assessment: Test at a glance*. Princeton, NJ: Author.
- Hambleton, R. K., & Pitoniak, M.J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 433-470). Westport, CT: American Council on Education/Praeger.
- Hambleton, R. K., & Plake, B.S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41-55.

Table 1. Committee Member Demographics

	Panel 1		Panel 2	
	N	Percent	N	Percent
Group you are representing				
Principal	10	43%	12	52%
Assistant Principal	2	9%	4	17%
College Faculty	8	35%	5	22%
Other	3	13%	2	9%
Race				
African American or Black	7	30%	8	35%
Alaskan Native or American Indian	0	0%	1	4%
Asian or Asian American	1	4%	1	4%
Native Hawaiian or Other Pacific Islander	0	0%	0	0%
White	14	61%	11	48%
Hispanic	1	4%	2	9%
Gender				
Female	11	48%	13	57%
Male	12	52%	10	43%
Are you certified as a school leader in your state?				
No	2	9%	2	9%
Yes	21	91%	21	91%
Are you currently a school leader in your state?				
No	6	26%	7	30%
Yes	17	74%	16	70%
Are you currently mentoring another school leader?				
No	10	43%	12	52%
Yes	13	57%	11	48%
How many years of experience do you have as a school leader in your state?				
3 years or less	2	9%	1	4%
4 - 7 years	4	17%	4	17%
8 - 11 years	5	22%	10	43%
12 - 15 years	3	13%	2	9%
16 years or more	9	39%	6	26%

Table 1. Committee Member Demographics

	Panel 1		Panel 2	
	N	Percent	N	Percent
For which education level are you currently a school leader?				
Elementary (<i>covering a combination of grades Pre-K to 6</i>)	5	22%	7	30%
Middle School (<i>covering a combination of grades 5 to 8</i>)	2	9%	2	9%
High School (<i>covering a combination of grades 9 to 12</i>)	5	22%	6	26%
Elementary & Middle School Combined	1	4%	1	4%
All Grades	1	4%	1	4%
Central Office	2	9%	2	9%
Higher Education	7	30%	4	17%
School Setting				
Urban	5	22%	10	43%
Suburban	8	35%	5	22%
Rural	10	43%	8	35%

Table 2A. Section I: Summary of Each Round of Judgments — Panel 1

Panelist	Round 1 Cutscore	Round 2 Cutscore	Round 3 Cutscore
1	55.05	54.55	54.25
2	60.95	58.25	57.50
3	46.80	48.45	51.35
4	46.70	47.30	47.70
5	52.90	51.10	50.90
6	46.95	51.95	52.25
7	52.90	53.00	53.30
8	50.40	50.10	50.40
9	53.50	51.15	51.15
10	36.90	36.90	36.90
11	53.50	53.45	54.65
12	57.30	58.75	59.25
13	45.10	45.10	45.70
14	52.50	55.00	55.55
15	45.90	45.55	44.95
16	47.90	48.10	48.90
17	45.40	45.30	45.50
18	48.40	47.80	48.30
19	54.90	54.60	55.30
20	40.50	41.25	42.45
21	55.30	55.35	54.55
22	49.85	49.55	48.50
23	48.50	45.80	45.80
Average	49.92	49.93	50.22
SD	5.49	5.30	5.23
SEJ	1.14	1.10	1.09
Highest	60.95	58.75	59.25
Lowest	36.90	36.90	36.90

Table 2B. Section I: Summary of Each Round of Judgments — Panel 2

Panelist	Round 1 Cutscore	Round 2 Cutscore	Round 3 Cutscore
1	67.90	66.05	64.70
2	62.90	60.65	60.75
3	52.90	53.20	53.00
4	37.70	40.00	40.20
5	49.75	50.55	50.75
6	42.50	43.10	43.20
7	58.95	58.15	58.45
8	53.50	51.95	51.95
9	48.20	46.55	45.75
10	55.05	54.85	54.65
11	51.20	49.60	48.70
12	68.75	68.00	68.60
13	43.25	44.30	45.80
14	51.95	52.75	53.35
15	43.70	45.50	46.05
16	54.70	52.35	52.25
17	55.10	55.40	55.70
18	45.35	47.25	47.95
19	62.70	63.35	63.55
20	43.25	44.75	45.75
21	59.60	60.50	60.95
22	45.60	47.10	48.00
23	59.30	58.60	58.60
Average	52.77	52.80	52.98
SD	8.43	7.62	7.47
SEJ	1.76	1.59	1.56
Highest	68.75	68.00	68.60
Lowest	37.70	40.00	40.20

Table 3A. Section II: Summary of Each Round of Judgments — Panel 1

Member	Round 1		Round 2		Round 3	
	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore
1	30.00	24.57	30.00	24.57	30.00	24.57
2	30.00	24.57	30.00	24.57	30.00	24.57
3	33.00	26.86	31.00	25.14	31.00	25.14
4	27.00	21.90	28.00	22.86	29.00	23.62
5	30.00	24.57	30.00	24.57	30.00	24.57
6	31.00	24.95	32.00	25.90	32.00	25.90
7	30.00	24.38	31.00	25.33	31.00	25.33
8	35.00	28.38	32.00	26.10	32.00	26.10
9	29.00	23.81	29.00	23.81	29.00	23.81
10	28.00	23.24	30.00	24.76	30.00	24.76
11	27.00	21.90	28.00	22.86	28.00	22.86
12	26.00	21.14	29.00	24.00	29.00	24.00
13	22.00	17.71	23.00	18.67	24.00	19.43
14	32.00	26.48	37.00	30.48	37.00	30.48
15	28.00	22.67	28.00	22.67	28.00	22.67
16	27.00	22.10	27.00	22.10	27.00	22.10
17	25.00	20.38	27.00	22.29	27.00	22.29
18	30.00	24.57	30.00	24.57	30.00	24.57
19	30.00	24.57	30.00	24.57	30.00	24.57
20	31.00	25.14	33.00	27.05	33.00	27.05
21	32.00	26.10	31.00	25.33	31.00	25.33
22	26.00	21.33	26.00	21.33	26.00	21.33
23	26.00	21.33	26.00	21.33	26.00	21.33
Average	28.91	23.59	29.48	24.12	29.57	24.19
SD	2.94	2.41	2.83	2.33	2.71	2.24
SEJ	0.61	0.50	0.59	0.49	0.57	0.47
Highest	35.00	28.38	37.00	30.48	37.00	30.48
Lowest	22.00	17.71	23.00	18.67	24.00	19.43

Note: The maximum raw score for Section II is 42 points, with a maximum weighted score of 34.27.

Table 3B. Section II: Summary of Each Round of Judgments — Panel 2

Member	Round 1		Round 2		Round 3	
	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore	Raw Score	Wt. Cutscore
1	31.00	25.14	29.00	23.62	29.00	23.62
2	34.00	27.81	32.00	25.90	32.00	25.90
3	30.00	24.38	29.00	23.62	29.00	23.62
4	20.00	16.19	24.00	19.43	25.00	20.19
5	25.00	20.76	25.00	20.76	26.00	21.52
6	30.00	24.38	29.00	23.62	29.00	23.62
7	29.00	23.43	30.00	24.38	30.00	24.38
8	32.00	25.90	32.00	26.29	32.00	26.29
9	31.00	25.14	27.00	21.90	28.00	23.05
10	31.00	25.14	30.00	24.38	30.00	24.38
11	27.00	22.29	28.00	23.05	28.00	23.05
12	31.00	25.52	31.00	25.52	31.00	25.52
13	31.00	25.33	31.00	25.33	31.00	25.33
14	29.00	23.62	30.00	24.38	30.00	24.38
15	26.00	21.52	22.00	18.48	22.00	18.48
16	29.00	23.81	21.00	17.33	26.00	21.52
17	29.00	23.81	30.00	24.57	30.00	24.57
18	27.00	22.29	29.00	23.81	30.00	24.57
19	18.00	14.48	26.00	21.14	26.00	21.14
20	23.00	19.05	26.00	21.33	26.00	21.33
21	23.00	18.48	23.00	18.48	24.00	19.24
22	23.00	19.05	26.00	21.33	27.00	22.10
23	31.00	25.33	31.00	25.33	31.00	25.33
Average	27.83	22.73	27.87	22.78	28.35	23.18
SD	4.11	3.34	3.21	2.58	2.67	2.15
SEJ	0.86	0.70	0.67	0.54	0.56	0.45
Highest	34.00	27.81	32.00	26.29	32.00	26.29
Lowest	18.00	14.48	21.00	17.33	22.00	18.48

Note: The maximum raw score for Section II is 42 points, with a maximum weighted score of 34.27.

Table 4A. Recommended SLLA Cutscores: Summary of Each Round of Judgments — Panel 1

Panelist	Round 1	Round 2	Round 3
	Overall Cutscore	Overall Cutscore	Overall Cutscore
1	79.62	79.12	78.82
2	85.52	82.82	82.07
3	73.66	73.59	76.49
4	68.60	70.16	71.32
5	77.47	75.67	75.47
6	71.90	77.85	78.15
7	77.28	78.33	78.63
8	78.78	76.20	76.50
9	77.31	74.96	74.96
10	60.14	61.66	61.66
11	75.40	76.31	77.51
12	78.44	82.75	83.25
13	62.81	63.77	65.13
14	78.98	85.48	86.03
15	68.57	68.22	67.62
16	70.00	70.20	71.00
17	65.78	67.59	67.79
18	72.97	72.37	72.87
19	79.47	79.17	79.87
20	65.64	68.30	69.50
21	81.40	80.68	79.88
22	71.18	70.88	69.83
23	69.83	67.13	67.13
Average	73.51	74.05	74.41
SD	6.42	6.34	6.26
SEJ	1.34	1.32	1.31
Highest	85.52	85.48	86.03
Lowest	60.14	61.66	61.66

Table 4B. Recommended SLLA Cutscores: Summary of Each Round of Judgments — Panel 2

Panelist	Round 1	Round 2	Round 3
	Overall Cutscore	Overall Cutscore	Overall Cutscore
1	93.04	89.67	88.32
2	90.71	86.55	86.65
3	77.28	76.82	76.62
4	53.89	59.43	60.39
5	70.51	71.31	72.27
6	66.88	66.72	66.82
7	82.38	82.53	82.83
8	79.40	78.24	78.24
9	73.34	68.45	68.80
10	80.19	79.23	79.03
11	73.49	72.65	71.75
12	94.27	93.52	94.12
13	68.58	69.63	71.13
14	75.57	77.13	77.73
15	65.22	63.98	64.53
16	78.51	69.68	73.77
17	78.91	79.97	80.27
18	67.64	71.06	72.52
19	77.18	84.49	84.69
20	62.30	66.08	67.08
21	78.08	78.98	80.19
22	64.65	68.43	70.10
23	84.63	83.93	83.93
Average	75.51	75.59	76.16
SD	9.94	8.73	8.34
SEJ	2.07	1.82	1.74
Highest	94.27	93.52	94.12
Lowest	53.89	59.43	60.39

Table 5A. Cutscores within 1 and 2 SEMs of the Recommended Cutscore — Panel 1

Recommended Cutscore (SEM)		Scale Score Equivalent
75 (5.05)		162
- 2 SEMs	65	151
-1 SEM	70	156
+1 SEM	81	169
+ 2 SEMs	86	175

Note: Consistent with the recommended cutscore, the cutscores at the different SEMs have been rounded to the next highest whole number.

Table 5B. Cutscores within 1 and 2 SEMs of the Recommended Cutscore — Panel 2

Recommended Cutscore (SEM)		Scale Score Equivalent
77 (4.98)		164
- 2 SEMs	68	154
-1 SEM	73	160
+1 SEM	82	170
+ 2 SEMs	87	176

Note: Consistent with the recommended cutscore, the cutscores at the different SEMs have been rounded to the next highest whole number.

Table 6A. SLLA Specification Judgments — Panel 1

	Very Important		Important		Slightly Important		Not Important	
	N	%	N	%	N	%	N	%
I. Vision and Goals (20%)	22	96%	1	4%	0	0%	0	0%
A. Vision and goals for teaching and learning	21	91%	2	9%	0	0%	0	0%
B. Shared commitments to implement the vision and goals	18	78%	5	22%	0	0%	0	0%
C. Continuous improvement toward the vision and goals	17	74%	6	26%	0	0%	0	0%
II. Teaching and Learning (30%)	23	100%	0	0%	0	0%	0	0%
A. Building a professional culture	20	87%	3	13%	0	0%	0	0%
B. Rigorous curriculum and instruction	19	83%	4	17%	0	0%	0	0%
C. Assessment and accountability	18	78%	5	22%	0	0%	0	0%
III. Managing Organizational Systems and Safety (10%)	11	48%	11	48%	1	4%	0	0%
A. Managing Operational Systems	6	26%	15	65%	2	9%	0	0%
B. Aligning and obtaining fiscal and human resources	11	48%	10	43%	2	9%	0	0%
C. Protecting the welfare and safety of students and staff	21	91%	2	9%	0	0%	0	0%
IV. Collaborating with Key Stakeholders (15%)	15	65%	8	35%	0	0%	0	0%
A. Collaborate with families and other community members	9	39%	14	61%	0	0%	0	0%
B. Community interests and needs	4	17%	17	74%	2	9%	0	0%
C. Maximizing community resources	4	17%	16	70%	3	13%	0	0%
V. Ethics and Integrity (15%)	23	100%	0	0%	0	0%	0	0%
A. Ethical and legal behavior	23	100%	0	0%	0	0%	0	0%
B. Personal values and beliefs	19	83%	4	17%	0	0%	0	0%
C. High standards for self and others	21	91%	2	9%	0	0%	0	0%
VI. The Education System (10%)	7	30%	15	65%	1	4%	0	0%
A. Professional influence	5	22%	17	74%	1	4%	0	0%
B. Managing local decisions within the larger educational policy environment	5	22%	15	65%	3	13%	0	0%

Table 6B. SLLA Specification Judgments — Panel 2

	Very Important		Important		Slightly Important		Not Important	
	N	%	N	%	N	%	N	%
I. Vision and Goals (20%)	19	83%	4	17%	0	0%	0	0%
A. Vision and goals for teaching and learning	18	78%	5	22%	0	0%	0	0%
B. Shared commitments to implement the vision and goals	16	70%	6	26%	1	4%	0	0%
C. Continuous improvement toward the vision and goals	17	74%	6	26%	0	0%	0	0%
II. Teaching and Learning (30%)	23	100%	0	0%	0	0%	0	0%
A. Building a professional culture	19	83%	4	17%	0	0%	0	0%
B. Rigorous curriculum and instruction	19	83%	4	17%	0	0%	0	0%
C. Assessment and accountability	17	74%	6	26%	0	0%	0	0%
III. Managing Organizational Systems and Safety (10%)	16	70%	7	30%	0	0%	0	0%
A. Managing Operational Systems	9	39%	13	57%	1	4%	0	0%
B. Aligning and obtaining fiscal and human resources	14	61%	9	39%	0	0%	0	0%
C. Protecting the welfare and safety of students and staff	23	100%	0	0%	0	0%	0	0%
IV. Collaborating with Key Stakeholders (15%)	15	65%	7	30%	1	4%	0	0%
A. Collaborate with families and other community members	16	70%	7	30%	0	0%	0	0%
B. Community interests and needs	5	22%	17	74%	1	4%	0	0%
C. Maximizing community resources	7	30%	14	61%	2	9%	0	0%
V. Ethics and Integrity (15%)	22	96%	1	4%	0	0%	0	0%
A. Ethical and legal behavior	23	100%	0	0%	0	0%	0	0%
B. Personal values and beliefs	17	74%	6	26%	0	0%	0	0%
C. High standards for self and others	21	91%	2	9%	0	0%	0	0%
VI. The Education System (10%)	7	30%	15	65%	1	4%	0	0%
A. Professional influence	8	35%	14	61%	1	4%	0	0%
B. Managing local decisions within the larger educational policy environment	5	22%	16	70%	1	4%	1	4%

Table 7A. Final Evaluation — Panel 1

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	%	N	%	N	%	N	%
I understood the purpose of this study.	23	100%	0	0%	0	0%	0	0%
The instructions and explanations provided by the facilitator were clear.	21	91%	2	9%	0	0%	0	0%
The training in the standard setting methods was adequate to give me the information I needed to complete my assignment.	20	87%	3	13%	0	0%	0	0%
The explanation of how the recommended cut scores are computed was clear.	18	78%	5	22%	0	0%	0	0%
The opportunity for feedback and discussion between rounds was helpful.	19	83%	4	17%	0	0%	0	0%
The process of making the standard setting judgments was easy to follow.	18	78%	5	22%	0	0%	0	0%

	Very Influential		Somewhat Influential		Not Influential	
	N	%	N	%	N	%
How influential was each of the following factors in guiding your standard setting judgments?						
The definition of the Just Qualified Candidate	17	74%	6	26%	0	0%
The between-round discussions	14	61%	9	39%	0	0%
The cut scores of other panel members	9	39%	11	48%	3	13%
My own professional experience	19	83%	4	17%	0	0%

	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	%	N	%	N	%	N	%
Overall, how comfortable are you with the panel's recommended cut score?	9	39%	9	39%	5	22%	0	0%

	Too Low		About Right		Too High	
	N	%	N	%	N	%
Overall, the panel's recommended cut score for the SLLA test is:	9	39%	13	57%	1	4%

Table 7B. Final Evaluation — Panel 2

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	%	N	%	N	%	N	%
I understood the purpose of this study.	23	100%	0	0%	0	0%	0	0%
The instructions and explanations provided by the facilitator were clear.	17	74%	6	26%	0	0%	0	0%
The training in the standard setting methods was adequate to give me the information I needed to complete my assignment.	22	96%	1	4%	0	0%	0	0%
The explanation of how the recommended cut scores are computed was clear.	18	78%	5	22%	0	0%	0	0%
The opportunity for feedback and discussion between rounds was helpful.	19	83%	4	17%	0	0%	0	0%
The process of making the standard setting judgments was easy to follow.	17	74%	5	22%	1	4%	0	0%

	Very Influential		Somewhat Influential		Not Influential	
	N	%	N	%	N	%
How influential was each of the following factors in guiding your standard setting judgments?						
The definition of the Just Qualified Candidate	20	87%	3	13%	0	0%
The between-round discussions	11	48%	12	52%	0	0%
The cut scores of other panel members	5	22%	14	61%	4	17%
My own professional experience	16	70%	7	30%	0	0%

	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	%	N	%	N	%	N	%
Overall, how comfortable are you with the panel's recommended cut score?	15	65%	8	35%	0	0%	0	0%

	Too Low		About Right		Too High	
	N	%	N	%	N	%
Overall, the panel's recommended cut score for the SLLA test is:	0	0%	20	87%	3	13%

Appendix A

Definition of a Just Qualified Candidate

1. Knows and uses the basic tenets of problem solving and conflict resolution
2. Collects, analyzes, and synthesizes data for instructional purposes and school processes to make decisions
3. Collaboratively develops and implements a shared vision and mission
4. Supervises, monitors, and evaluates the impact of the instructional program
5. Manages resources in an effective, efficient, and leveraging manner (including financial , human, and physical resources)
6. Knows how to create and maintain a safe school environment
7. Builds and sustains positive relationships with all stakeholders through effective communication and collaboration
8. Able to lead a diverse student population and to meet the needs of all students
9. Models principles of self-awareness, reflective practice *and* transparency ~~and loyalty to a shared vision~~⁵
10. Advocates for children, families and caregivers
11. Understands applicable local, state, and federal laws and guidelines as they affect student learning
12. Acts with integrity, fairness and in an ethical manner toward all stakeholders
13. Knows and *applies* ~~uses~~ the basic tenets of teaching and learning (including the use of technology to support teaching and learning) *as it impacts student achievement*⁵

⁵ Revisions to the JQC definition based on discussions during Panel 2 are indicated.

Appendix B

List of Panelists – Panel 1

David W. Adams	Hinds County School District-Utica Elementary/ Middle School (Missouri)
Martin Dickerson	W.O. Krumbiegel Middle School (New Jersey)
Willie S. Dickerson	Williamson County Board of Education (Tennessee)
Maureen Fitzpatrick	Sacred Heart University (Connecticut)
Ann R. Hardy	Vermilion Parish School District (Louisiana)
Patrick Hartnett	Leavitt Area High School (Maine)
Denise Harwood	Eldon R-I Schools (Missouri)
Hamlet M. Hernandez	Hamdew Public Schools (Connecticut)
Ruthanne A. Keller	Davis School District (Utah)
Robert Lyons	Murray State University (Kentucky)
Shirley Marie McCarther	University of Missouri-Kansas City
Suzanne S. McCotter	Montclair State University (New Jersey)
Marjorie E. Miles	Coppin State University (Maryland)
David Lee Parker	Baltimore County Public Schools (Maryland)
Tony Pellegrini	Southern Utah University
Shawn Pelote	D.C. Public Schools (Washington DC)
Perry L. Perkins	University of Kansas
Bridget Thomas	Lake Arthur High School (Louisiana)
David Treick	Cody High School (Wyoming)
Sheila Weathersby-Burbridge	Columbia School District-Columbia High School (Mississippi)
Debra L. Williams	Fayette County School Corporation (Indiana)
R. Kieth Williams	Harding University (Arkansas)
Wayne Yamagishi	Association of California School Administrators

List of Panelists – Panel 2

Matthew Boggan	Mississippi State University-Educational Leadership Department
Sheila K. Brown	Old Saybrook Public Schools (Connecticut)
Adrian Charley	D.C. Public Schools (Washington D.C.)
Augusta A. Clark	University of Louisiana-Monroe
Troy Clawson	South Callaway High School (Missouri)
Sharonica L. Hardin	St. Louis Public School District (Missouri)
Donnie Renée Johnson	The Schomburg Charter School (North Carolina)
James H. Kirk Jr.	Alcoa City Schools-Alcoa Middle School (Tennessee)
Sharon W. Lair	West Baton Rouge Parish School System (Louisiana)
Renata S. Lantos	Bielefield School-Middletown Public Schools (Connecticut)
William O. Lawson, Jr.	Hinds County School District (Mississippi)
Debbi P. Lindsey	Caverna Independent Schools (Kentucky)
Gary McGuire	Point Loma Nazarene University (California)
Mark C. Mitchell	Alta High School (Utah)
Robert A. Motley	Howard County Public Schools (Maryland)
Bobby I. Occena	Bertie Early College High School (North Carolina)
Tom Pyron	Fulton County Area Technology Center (Kentucky)
Deborah F. Sharpe	Baltimore City Public School System (Maryland)
Elizabeth Vaughn-Neely	University of Arkansas at Little Rock
Kevin J. Walsh	William Patterson University of New Jersey
Jane E. White-Kilcollins	Hilltop Elementary School-Caribou School Department (Maine)
Sylvia H. Wilkins	Dillard Drive Elementary School-Wake County Public Schools (North Carolina)
Linda Wolfskill	Moorcroft Elementary School (Wyoming)

Appendix C Agenda School Leaders Licensure Assessment Standard Setting Study

DAY 1

- 8:00 – 8:15 AM Welcome and Introductions
- 8:15 – 9:00 Overview of Standard Setting and Workshop Events
Sign nondisclosure and complete biographical information form
- 9:00 – 11:00 Review of the SLLA
Break as needed
- 11:00 – 11:30 Discuss SLLA
What is being measured?
What does an entering school (education) leader need to know and do?
- 11:30 – 12:15 PM Define Knowledge/Skills of Just Qualified Candidate
- 12:15 – 1:00 Lunch
- 1:00 – 1:30 Define Knowledge/Skills of Just Qualified Candidate (continued)
- 1:30 – 2:15 Standard Setting Training for Multiple-Choice Questions
Practice judgments -- first 3 questions
Complete training evaluation form
- 2:15 – 3:45 Round 1 Standard Setting Judgments for Multiple-Choice Questions
- 3:45 – 4:15 Standard Setting Training for Constructed-Response Questions
Practice judgments – first question
Complete training evaluation form
- 4:15 – 4:45 Round 1 Standard Setting Judgments for Constructed-Response Questions
- 4:45 – 5:00 Collect Materials; End of Day 1

AGENDA
School Leaders Licensure Assessment
Standard Setting

DAY 2

- 8:00 – 8:15 AM Questions From Day 1 and Overview of Day 2
- 8:15 – 10:40 Round 1 Feedback and Discussion
- 10:40 – 11:00 Round 2 Standard Setting Judgments MC and CR Questions
- 11:00 – 11:15 Data Entry; Break
- 11:15 – 12:00 PM Round 2 Feedback and Discussion
- 12:00 – 12:45 Lunch
- 12:45 – 1:00 Round 3 Standard Setting Judgments MC and CR Questions
- 1:00 – 1:30 Specification Judgment Training
Practice judgments – first specification: *Vision and Goals*
- 1:30 – 1:45 Complete Specification Judgments
- 1:45 – 2:00 Feedback on Round 3 Recommended Cut Score
- 2:00 – 2:15 Complete Final Evaluation
- 2:15 – 2:30 Collect Materials; End of Study

APPENDIX C

School Leaders Licensure Assessment (SLLA) Test at a Glance

School Leaders Licensure Assessment (1011)

<i>Test at a Glance</i>				
Test Name	School Leaders Licensure Assessment			
Test Code	1011			
Time	4 hours			
Number of Questions	7 constructed-response questions and 100 multiple-choice questions			
	Content Categories	Percentage of Examination	Multiple-choice questions (approx.)	Constructed response questions
	I. Vision and Goals	20%	18	2
	II. Teaching and Learning	30%	25	3
	III. Managing Organizational Systems and Safety	10%	15	–
	IV. Collaborating with Key Stakeholders	15%	21	–
	V. Ethics and Integrity	15%	21	–
	VI. The Education System	10%	–	2

About This Test

The School Leaders Licensure Assessment measures whether entry-level education leaders have the standards-relevant knowledge believed necessary for competent professional practice. The content of the assessment was defined by a National Advisory Committee of expert practitioners and preparation faculty and confirmed by a national survey of the field. The content is aligned with the *Educational Leadership Policy Standards: ISLLC 2008*.

The four-hour assessment is divided into two separately timed sections:

Section I (2 hours 20 minutes) – 100 multiple-choice questions.

Section II (1 hour 40 minutes) – Seven constructed-response questions calling for written answers based on scenarios and sets of documents that an education leader might encounter. Each of the seven constructed-response questions in the test focuses on a specific content area related to the standards addressed in *ISLLC 2008*. In answering the questions, candidates are required to analyze situations and data, to propose appropriate courses of action, and to provide rationales for their proposals.

APPENDIX D

*School Leaders Licensure Assessment (SLLA)
State Passing Scores (as of November 9, 2009)*

Revised School Leaders Licensure Assessment (SLLA)

State Passing Scores (as of November 9, 2009)

<u>State</u>	<u>Score</u>
Arkansas	No Score Set Yet
Connecticut	No Score Set Yet
District of Columbia	163
Indiana	163
Kansas	165
Kentucky	No Score Set Yet
Maine	No Score Set Yet
Maryland	No Score Set Yet
Mississippi	169
Missouri	163
New Jersey	163
North Carolina	163
Tennessee	No Score Set Yet
Utah	163
Virgin Islands	156

APPENDIX E

*School Leaders Licensure Assessment (SLLA)
Scaled Score Distribution for Entire Sample
For SLLA2 – September 2009 Administration*

***Scaled Score Distribution for Entire Sample
For SLLA2 -- September 2009 Administration***

Scaled Score	Frequency	Cumulative Percent
196	1	0.09
195	1	0.17
194	1	0.26
193	4	0.61
192	5	1.04
190	3	1.30
189	11	2.25
188	9	3.03
187	15	4.33
186	18	5.89
185	22	7.80
184	26	10.05
183	42	13.69
181	42	17.33
180	43	21.06
179	41	24.61
178	42	28.25
177	55	33.02
176	42	36.66
175	52	41.16
174	47	45.23
172	43	48.96
171	60	54.16
170	45	58.06
169	38	61.35
168	41	64.90
167	46	68.89
166	43	72.62
165	27	74.96
164	33	77.82
162	34	80.76
161	24	82.84
160	28	85.27
159	18	86.83
158	22	88.73
157	19	90.38

***Scaled Score Distribution for Entire Sample
For SLA2 -- September 2009 Administration***

Scaled Score	Frequency	Cumulative Percent
156	19	92.03
155	16	93.41
153	15	94.71
152	10	95.58
151	4	95.93
150	6	96.45
149	4	96.79
148	3	97.05
147	3	97.31
146	4	97.66
144	3	97.92
143	2	98.09
142	5	98.53
141	3	98.79
140	4	99.13
139	3	99.39
138	1	99.48
137	1	99.57
133	2	99.74
132	1	99.83
131	1	99.91
123	1	100.00

*Scaled Score Distribution for Entire Sample
For SLLA2 -- September 2009 Administration*

The MEANS Procedure

Analysis Variable : Scaled Score				
N	Mean	Std Dev	Minimum	Maximum
1154	170.69	10.73	123.00	196.00