



Multistate Standard-Setting Technical Report

**ETS<sup>®</sup> SCHOOL LEADER LICENSURE ASSESSMENT (6990)**

Educational Testing Service

Princeton, New Jersey

February 2018

---

# EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *ETS*<sup>®</sup> School Leader Licensure Assessment (SLLA), research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

## PARTICIPATING STATES

Panelists from 20 states and Washington, DC were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either school leaders or college faculty who prepare school leaders and (b) familiarity with the knowledge and skills required of beginning school leaders.

## RECOMMENDED PASSING SCORE

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the SLLA, the recommended passing score<sup>1</sup> is 77 out of a possible 133 raw-score points. The scale score associated with a raw score of 77 is 151 on a 100–200 scale.

---

<sup>1</sup> Results from the two panels participating in the study were averaged to produce the recommended passing score.

To support the decision-making process for education agencies establishing a passing score (cut score) for the *ETS*<sup>®</sup> School Leader Licensure Assessment (SLLA), research staff from ETS designed and conducted a multistate standard-setting study in January 2018 in Princeton, New Jersey. Education agencies<sup>2</sup> recommended panelists with (a) experience as either school leaders or college faculty who prepare school leaders and (b) familiarity with the knowledge and skills required of beginning school leaders. Twenty states and Washington, DC (Table 1) were represented by 34 panelists. (See Appendix A for the names and affiliations of the panelists.)

**Table 1**  
***Participating Jurisdictions and Number of Panelists***

---

Alabama (2 panelists)	New Jersey (1 panelist)
Arkansas (2 panelists)	North Dakota (2 panelists)
Connecticut (2 panelists)	Pennsylvania (1 panelist)
Delaware (1 panelist)	Rhode Island (1 panelists)
Hawaii (1 panelist)	South Dakota (1 panelist)
Idaho (1 panelist)	Tennessee (2 panelists)
Kansas (2 panelists)	Utah (2 panelists)
Kentucky (2 panelists)	Virginia (3 panelists)
Maryland (1 panelist)	Washington, DC (2 panelists)
Mississippi (2 panelists)	West Virginia (1 panelist)
Nebraska (2 panelists)	

---

The following technical report contains three sections. The first section describes the content and format of the test. The second section describes the standard-setting processes and methods. The third section presents the results of the standard-setting study.

ETS provides a recommended passing score from the multistate standard-setting study to education agencies. In each jurisdiction, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in accordance with applicable regulations. This study provides a recommended passing score,<sup>3</sup> which

---

<sup>2</sup> States and jurisdictions that currently use any ETS educator licensure test were invited to participate in the multistate standard-setting study.

<sup>3</sup> In addition to the recommended passing score averaged across the two panels, the recommended passing scores for each panel are presented.

represents the combined judgments of two panels of experienced educators. Each jurisdiction may want to consider the recommended passing score but also other sources of information when setting the final SLLA passing score (see Geisinger & McCormick, 2010). A jurisdiction may accept the recommended passing score, adjust the score upward to reflect more stringent expectations, or adjust the score downward to reflect more lenient expectations. There is no *correct* decision; the appropriateness of any adjustment may only be evaluated in terms of its meeting the jurisdiction's needs.

Two sources of information to consider when setting the passing score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the SLLA score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a jurisdiction to recognize that any test score on any standardized test—including a SLLA score—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows a jurisdiction to gauge the likelihood that the recommended passing score from a particular panel would be similar to the passing scores recommended by other panels of experts similar in composition and experience. The smaller the SEJ, the more likely that another panel would recommend a passing score consistent with the recommended passing score. The larger the SEJ, the less likely the recommended passing score would be reproduced by another panel.

In addition to measurement error metrics (e.g., SEM, SEJ), each jurisdiction should consider the likelihood of classification errors. That is, when adjusting a passing score, policymakers should consider whether it is more important to minimize a false-positive decision or to minimize a false-negative decision. A false-positive decision occurs when a candidate's test score suggests that he should receive a license/certificate, but his actual level of knowledge/skills indicates otherwise (i.e., the candidate does not possess the required knowledge/skills). A false-negative decision occurs when a candidate's test score suggests that she should not receive a license/certificate, but she actually does possess the required knowledge/skills. The jurisdiction needs to consider which decision error is more important to minimize.

---

# OVERVIEW OF THE *ETS*<sup>®</sup> SCHOOL LEADER LICENSURE ASSESSMENT

The *ETS*<sup>®</sup> *School Leadership Series* Study Companion for the School Leader Licensure Assessment (ETS, in press) describes the purpose and structure of the test. In brief, the test measures the extent to which entry-level school leaders demonstrate the standards-relevant knowledge and skills necessary for competent professional practice. The test is aligned to the National Policy Board for Educational Administration (NPBEA) *Professional Standards for Educational Leaders* (NPBEA, 2015) and the draft *National Educational Leadership Preparation* (NELP) building-level standards (UCEA, 2016).

The four-hour assessment contains 120 selected-response items<sup>4</sup> and four constructed-response items covering seven content areas: *Strategic Leadership* (approximately 20 selected-response items), *Instructional Leadership* (approximately 27 selected-response items), *Climate and Cultural Leadership* (approximately 22 selected-response items), *Ethical Leadership* (approximately 19 selected-response items), *Organizational Leadership* (approximately 16 selected-response items), *Community Engagement Leadership* (approximately 16 selected-response items) and *Analysis* (4 constructed-response items).<sup>5</sup> The reporting scale for the SLLA ranges from 100 to 200 scale-score points.

---

## PROCESSES AND METHODS

The design of the standard-setting study included two expert panels. Before the study, panelists received an email explaining the purpose of the standard-setting study and requesting that they review the content specifications for the test. This review helped familiarize the panelists with the general structure and content of the test.

For each panel, the standard-setting study began with a welcome and introduction by the meeting facilitator. The facilitator described the test, provided an overview of standard setting, and presented the agenda for the study. Appendix B shows the agenda for the panel meeting.

---

<sup>4</sup> Twenty of the 120 selected-response items are pretest items and do not contribute to a candidate's score.

<sup>5</sup> The number of selected-response items for each content area may vary slightly from form to form of the test.

## REVIEWING THE TEST

The standard-setting panelists first took the test and then discussed it. This discussion helped bring the panelists to a shared understanding of what the test does and does not cover, which serves to reduce potential judgment errors later in the standard-setting process.

The test discussion covered the major content areas being addressed by the test. Panelists were asked to remark on any content areas that would be particularly challenging for entry-level school leaders or areas that address content particularly important for entry-level school leaders.

## DEFINING THE JUST QUALIFIED CANDIDATE

Following the review of the test, panelists described the just qualified candidate. The *just qualified candidate description* plays a central role in standard setting (Perie, 2008); the goal of the standard-setting process is to identify the test score that aligns with this description.

Both panels worked together to create a description of the just qualified candidate — the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, they first split into smaller groups to consider the just qualified candidate. Then they reconvened and, through whole-group discussion, created the description of the just qualified candidate to use for the remainder of the study. After the description was completed, panelists were split into two, distinct panels that worked separately for the remainder of the study.

The written description of the just qualified candidate summarized the panel discussion in a bulleted format. The description was not intended to describe all the knowledge and skills of the just qualified candidate but only highlight those that differentiate a *just* qualified candidate from a *not quite* qualified candidate. The written description was distributed to panelists to use during later phases of the study (see Appendix C for the just qualified candidate description).

## PANELISTS' JUDGMENTS

The SLLA includes both dichotomously-scored (selected-response items) and constructed-response items. Panelists received training in two distinct standard-setting approaches: one standard-setting approach for the dichotomously-scored items and another approach for the constructed-response items.

A panel's passing score is the sum of the interim passing scores recommended by the panelists for (a) the dichotomously-scored items and (b) the constructed-response items. As with scoring and reporting, the panelists' judgments for the constructed-response items were weighted such that they contributed 25% of the overall score.

**Dichotomously scored items.** The standard-setting process for the dichotomously-scored items was a probability-based Modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). In this study, each panelist judged each item on the likelihood (probability or chance) that the just qualified candidate would answer the item 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 the just qualified candidate would answer the item correctly because the item is difficult for the just qualified candidate. The higher the value, the more likely it is that the just qualified candidate would answer the item correctly.

Panelists were asked to approach the judgment process in two stages. First, they reviewed both the description of the just qualified candidate and the item. Then the panelists estimated what chance a just qualified candidate would have of answering the question correctly. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Items in the 0 to .30 range were those the just qualified candidate would have a low chance of answering correctly.
- Items in the .40 to .60 range were those the just qualified candidate would have a moderate chance of answering correctly.
- Items in the .70 to 1 range were those that the just qualified candidate would have a high chance of answering correctly.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that there was a high chance that the just qualified candidate would answer the question correctly, the initial decision would be in the .70 to 1 range. The second decision for the panelist was to judge if the likelihood of answering it correctly is .70, .80, .90, .95 or 1.

After the training, panelists made practice judgments and discussed those judgments and their rationale. All panelists completed a post-training survey to confirm that they had received adequate

training and felt prepared to continue; the standard-setting process continued only if all panelists confirmed their readiness.

**Constructed-response items.** An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for the constructed-response items. For this portion of the study, a panelist decided on the assigned score value that would most likely be earned by the just qualified candidate for each constructed-response item. Panelists were asked first to review the definition of the just qualified candidate and then to review the constructed-response item and its rubric. The rubric for a constructed-response item defines (holistically) the quality of the evidence that would merit a response earning a particular score. During this review, each panelist independently considered the level of knowledge/skill required to respond to the constructed-response item and the features of a response that would earn a particular score, as defined by the rubric. Each panelist decided on the score most likely to be earned by the just qualified candidate from the possible values a test taker can earn.

A test-taker's response to a constructed-response item is independently scored by two raters, and the sum of the raters' scores is the assigned score<sup>6</sup>; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). For their ratings, each panelist decided on the score most likely to be earned by a just qualified candidate from the following possible values: 0, 1, 2, 3, 4, 5, or 6. For each of the constructed-response item, panelists recorded the score (0 through 6) that a just qualified candidate would most likely earn.

After the training, panelists made practice judgments and discussed those judgments and their rationale. All panelists completed a post-training survey to confirm that they had received adequate training and felt prepared to continue; the standard-setting process continued only if all panelists confirmed their readiness.

**Multiple Rounds.** Following this first round of judgments (*Round 1*), item-level feedback was provided to the panel. The panelists' judgments were displayed for each item and summarized across panelists. For dichotomously-scored items, items were highlighted to show when panelists converged in their judgments (at least two-thirds of the panelists located an item in the same difficulty range) or diverged in their judgments.

---

<sup>6</sup> If the two raters' scores differ by more than one point (non-adjacent), the Chief Reader for that item assigns the score, which is then doubled.



The panelists discussed their item-level judgments. These discussions helped panelists maintain a shared understanding of the knowledge/skills of the just qualified candidate and helped to clarify aspects of items that might not have been clear to all panelists during the Round 1 judgments. The purpose of the discussion was not to encourage panelists to conform to another's judgment, but to understand the different relevant perspectives among the panelists.

In Round 2, panelists discussed their Round 1 judgments and were encouraged by the facilitator (a) to share the rationales for their judgments and (b) to consider their judgments in light of the rationales provided by the other panelists. Panelists recorded their Round 2 judgments only for items when they wished to change a Round 1 judgment. Panelists' final judgments for the study, therefore, consist of their Round 1 judgments and any adjusted judgments made during Round 2.

Other than the description of the just qualified candidate, results from Panel 1 were not shared with Panel 2. The item-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

---

## RESULTS

### EXPERT PANELS

Table 2 presents a summary of the panelists' demographic information. The panel included 34 panelists representing 20 states and Washington, DC (See Appendix A for a listing of panelists.) Fourteen panelists were principals, two were vice principals, two were superintendents, one was a building-level instructional team leader, 13 were college faculty, and two were college administrators. All thirteen faculty members' job responsibilities included the training of school leaders.

The demographic information by panel is presented in Appendix D (Table D1).

**Table 2**  
*Panel Member Demographics (Across Panels)*

	<i>N</i>	<i>%</i>
<b>Current position</b>		
Principal	14	41
Vice Principal	2	6
Superintendent	2	6
Instructional Team Leader	1	3
College faculty	13	38
College Administrator	2	6
<b>Race</b>		
White	25	74
Black or African American	5	15
Asian or Asian American	1	3
American Indian or Alaskan Native	1	3
Other	2	6
<b>Gender</b>		
Female	17	50
Male	17	50
<b>Are you currently certified as a school leader in your state?</b>		
Yes	19	56
No	0	0
I am not a school leader	15	44

---

**Table 2 (continued)****Panel Member Demographics (Across Panels)**

	<i>N</i>	<i>%</i>
<b>Including this year, how many years of experience do you have as an educational leader?</b>		
3 years or less	1	3
4 - 7 years	6	18
8 - 11 years	6	18
12 - 15 years	4	12
16 years or more	2	6
I am not a school leader	15	44
<b>If you are building level school leader, what grade levels are taught in your school?</b>		
Elementary	8	24
Middle School	2	6
High School	7	21
I am not a school leader	17	50
<b>If you are building-level school leader, which best describes the location of your school?</b>		
Urban	4	12
Suburban	5	15
Rural	8	24
I am not a school leader	17	50
<b>Are you currently involved in the training or preparation of school leaders?</b>		
Yes	15	44
No	0	0
I am not college faculty	19	56
<b>How many years of experience (including this year) do you have preparing school leaders?</b>		
3 years or less	0	0
4 - 7 years	0	0
8 - 11 years	4	12
12 - 15 years	3	9
16 years or more	8	24
Not college faculty	19	56

## STANDARD-SETTING JUDGMENTS

Table 3 summarizes the standard-setting judgments (Round 2) of panelists. The table also includes estimates of the measurement error associated with the judgments: the standard deviation of the mean and the standard error of judgment (SEJ). The SEJ is one way of estimating the reliability or consistency of a panel’s standard-setting judgments.<sup>7</sup> It indicates how likely it would be for several other panels of educators similar in makeup, experience, and standard-setting training to the current panel to recommend the same passing score on the same form of the test. The confidence intervals created by adding/subtracting two SEJs to each panel’s recommended passing score overlap, indicating that they may be comparable.

Panelist-level results, for Rounds 1 and 2, are presented in Appendix D (Table D2).

**Table 3**  
*Summary of Round 2 Standard-setting Judgments*

	<b>Panel 1</b>	<b>Panel 2</b>
Average	76.58	76.58
Lowest	66.27	65.47
Highest	87.80	90.32
SD	6.19	6.87
SEJ	1.50	1.67

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 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 each panel (see Table D2 in Appendix D). The Round 2 average score is the panel’s recommended passing score.

The panels’ passing score recommendations for the SLLA are 76.58 for Panel 1 and 76.58 for Panel 2 (out of a possible 133 raw-score points). The values were rounded to the next highest whole number, to determine the functional recommended passing score — 77 for both Panels 1 and 2. The scale score associated with 77 raw points is 151.

---

<sup>7</sup> An SEJ assumes that panelists are randomly selected and that standard-setting judgments are independent. It is seldom the case that panelists are randomly sampled, and only the first round of judgments may be considered independent. The SEJ, therefore, likely underestimates the uncertainty of passing scores (Tannenbaum & Katz, 2013).

In addition to the recommended passing score for each panel, the average passing score across the two panels is provided to help education agencies determine an appropriate passing score. The panels' average passing score recommendation for the SLLA is 76.58 (out of a possible 133 raw-score points). The value was rounded to 77 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 77 raw points is 151.

Table 4 presents the estimated conditional standard error of measurement (CSEM) around the recommended passing score (the average across the two panels). A standard error represents the uncertainty associated with a test score. The scale scores associated with one and two CSEM above and below the recommended passing score are provided. The conditional standard error of measurement provided is an estimate.

**Table 4**  
*Passing Scores Within 1 and 2 CSEM of the Recommended Passing Score<sup>8</sup>*

<b>Recommended passing score (CSEM)</b>		<b>Scale score equivalent</b>
	77 (5.54)	151
-2 CSEM	66	140
-1 CSEM	72	146
+ 1 CSEM	83	157
+ 2 CSEM	89	163

**Note.** CSEM = conditional standard error(s) of measurement.

## FINAL EVALUATIONS

The panelists completed an evaluation at the conclusion of their standard-setting study. The evaluation asked the panelists to provide feedback about the quality of the standard-setting implementation and the factors that influenced their decisions. The responses to the evaluation provided evidence of the validity of the standard-setting process, and, as a result, evidence of the reasonableness of the recommended passing score.

Panelists were also shown their panel's recommended passing score and asked (a) how comfortable they are with the recommended passing score and (b) if they think the score was too high, too low, or about right. A summary of the final evaluation results is presented in Appendix D.

<sup>8</sup> The unrounded CSEM value is added to or subtracted from the rounded passing-score recommendation. The resulting values are rounded up to the next-highest whole number and the rounded values are converted to scale scores.

All panelists *strongly agreed* or *agreed* that they understood the purpose of the study; all but one *strongly agreed*. All panelists *strongly agreed* or *agreed* that the facilitator’s instructions and explanations were clear. All panelists *strongly agreed* or *agreed* that they were prepared to make their standard-setting judgments. All panelists *strongly agreed* or *agreed* that the standard-setting process was easy to follow.

All panelists reported that the description of the just qualified candidate was at least *somewhat influential* in guiding their standard-setting judgments; 24 of the 34 panelists indicated the description was *very influential*. All of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Two-thirds of the panelists (23 of the 34 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

All but one of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 27 of the 34 panelists were *very comfortable*. Thirty-two of the 34 panelists indicated the recommended passing score was *about right*; the remaining two panelists indicated that the passing score was *too low*.

---

## SUMMARY

To support the decision-making process for education agencies establishing a passing score (cut score) for the SLLA, research staff from ETS designed and conducted a multistate standard-setting study.

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the SLLA, the recommended passing score<sup>9</sup> is 77 out of a possible 133 raw-score points. The scale score associated with a raw score of 77 is 151 on a 100–200 scale.

---

<sup>9</sup> Results from the two panels participating in the study were averaged to produce the recommended passing score.

---

## 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.
- ETS. (in press). *The ETS® School Leadership Series: Study Companion: School Leader Licensure Assessment (6990)*. Princeton, NJ: Author.
- Geisinger, K. F. & McCormick, C. M. (2010), Adopting Cut Scores: Post-Standard-Setting Panel Considerations for Decision Makers. *Educational Measurement: Issues and Practice, 29*: 38–44.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th 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.
- National Policy Board for Educational Administration (2015). *Professional Standards for Educational Leaders 2015*. Reston, VA: Author.
- Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practice, 27*, 15–29.
- Tannenbaum, R. J., & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol. 3. Testing and assessment in school psychology and education* (pp. 455–477). Washington, DC: American Psychological Association.
- University Council for Educational Administration (2016). *Draft National Educational Leadership Preparation (NELP) Standards for Building Level Leaders*. Charlottesville, VA: Author.

---

## APPENDIX A

# PANELISTS' NAMES & AFFILIATIONS



***Participating Panelists With Affiliation***

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Sousan Arafeh	Southern Connecticut State University (CT)
Carrie Ballinger	Eastern Kentucky University (KY)
Jesse Boyd	King George County Schools (VA)
Patricia Brandom-Pride	D.C. Public Schools (DC)
Harrie Buecker	University of Louisville (KY)
Dennis Bunch	The University of Mississippi (MS)
John Burke	Haysville USD 261/Newman University (KS)
Kyley Cumbow	Georgia Morse Middle School, Pierre (SD)
Nicolle Currie	Rural Point Elementary School/Hanover County Public Schools (VA)
Lori DeSimone	North Providence School Department (RI)
Kevin DiCostanzo	Delaware Department of Education/Milford School District (DE)
Docia Generette	Shelby County Schools (TN)
Angela Goodloe	Norfolk State University (VA)
Lisa Grillo	Howard University School of Education (DC)
Clarence H. Horn	Fort Hays State University (KS)
Matt Kiser	Homewood City Schools, Edgewood Elementary School (AL)
Carmelita Lamb	University of Mary, Bismarck (ND)
James McIntyre	University of Tennessee (TN)
Justin S. N. Mew	Henry J. Kaiser High School (HI)
Amy Mitchell	Washington County School District (UT)
Janice Page Johnson	Greenville Public School District (MS)

---

*Participating Panelists With Affiliation (continued)*

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Craig Pease	Wayne Sate College (NE)
Christopher Pritchett	Troy University (AL)
Taylor Raney	University of Idaho (ID)
Christopher Rau	Regional School District #10 (CT)
Russ Riehl	Simle Middle School, Bismarck Public schools (ND)
Bess Scott	Doane University (NE)
Daniel Shea	Hood College (MD)
Mark Shumate	Greewood Public Schools (AR)
Stefanie Smithey	Carroll Smith Elementary School (AR)
Karen Soper	Manti Elementary School (UT)
Thomas Traver	Dallas School District (PA)
Eugenia Webb-Damron	Marshall University (WV)
Anthony C. Wright	Wilmington University (DE)

---

---

**APPENDIX B**  
**STUDY AGENDA**

# AGENDA

## ***ETS*<sup>®</sup> School Leader Licensure Assessment (SLLA) Standard-Setting Study**

### Day 1

Welcome and Introduction

Overview of Standard Setting and the SLLA

Review the SLLA

Discuss the SLLA

Define the Knowledge/Skills of a Just Qualified Candidate

Standard Setting Training for Selected-Response Items

Round 1 Judgments for Selected-Response Items

Collect Materials; End of Day 1

### Day 2

Overview of Day 2

Standard Setting Training for Constructed-Response Items

Round 1 Judgments for Constructed-Response Items

Round 1 Feedback and Round 2 Judgments

Feedback on Round 2 Recommended Cut Score

Complete Final Evaluation

Collect Materials; End of Study

---

## APPENDIX C

### JUST QUALIFIED CANDIDATE DESCRIPTION

## Description of the Just Qualified Candidate<sup>10</sup>

### A just qualified candidate ...

#### I. Strategic Leadership

1. Knows multiple sources are needed for data analysis to inform continuous improvement
2. Knows how local/state/federal policies impact school operations
3. Understands the value of engaging stakeholders with diverse perspectives
4. Knows that there is value in having and implementing a mission, a vision, goals and core values

#### II. Instructional Leadership

1. Familiar with how to use student/teacher data to drive differentiated professional development needs
2. Is familiar with the need for alignment of curriculum and instruction, student assessments, professional development, and reporting tools with content standards
3. Understands the use of valid assessments to improve instruction and student achievement

#### III. Climate and Cultural Leadership

1. Understands the importance of fostering a supportive, collaborative, respectful working environment
2. Understands the need for equitable access to learning opportunities
3. Understands the need to implement policies and procedures in a fair, unbiased, and culturally-responsive manner
4. Understands the need to create and sustain a school environment to meet the academic, emotional, social, and physical needs of students

#### IV. Ethical Leadership

1. Understands, models, and promotes integrity and ethical leadership
2. Knows how to maintain standards and accountability for ethical and legal behavior among faculty, staff and students

---

<sup>10</sup> Description of the just qualified candidate focuses on the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate.

## Description of the Just Qualified Candidate<sup>11</sup> (continued)

### A just qualified candidate ...

#### V. Organizational Leadership

1. Knows how to interpret and apply district policies to monitor and sustain the operation of the school
2. Is familiar with the allocation of fiscal and personnel resources to support students' needs
3. Knows how to develop and widely communicate a system of support for student welfare and safety

#### VI. Community Engagement Leadership

1. Understands the importance of engaging families in educational decision-making through two-way communication and collaborative partnerships
2. Is familiar with the need to solicit, identify, and value diverse perspectives
3. Knows the importance of developing mutually beneficial school-community relationships
4. Is familiar with how to seek community resources

#### VII. Analysis

1. Familiar with the need for a coherent, collaborative, and comprehensive school plan that will enable learning and success for all students

---

<sup>11</sup> Description of the just qualified candidate focuses on the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate.

---

# APPENDIX D

## RESULTS



**Table D1**  
**Panel Member Demographics (by Panel)**

	<b>Panel 1</b>		<b>Panel 2</b>	
	<i>N</i>	%	<i>N</i>	%
<b>Current position</b>				
Principal	8	47	6	35
Vice Principal	0	0	2	12
Superintendent	0	0	2	12
Instructional Team Leader	1	6	0	0
College Faculty	8	47	5	29
College Administrator	0	0	2	12
<b>Race</b>				
White	12	71	13	76
Black or African American	2	12	3	18
Asian or Asian American	1	6	0	0
American Indian or Alaskan Native	0	0	1	6
Other	2	12	0	0
<b>Gender</b>				
Female	8	47	9	53
Male	9	53	8	47
<b>Are you currently certified as a school leader in your state?</b>				
Yes	9	53	10	59
No	0	0	0	0
I am not a school leader	8	47	7	41
<b>Including this year, how many years of experience do you have as an educational leader?</b>				
3 years or less	0	0	1	6
4 - 7 years	3	18	3	18
8 - 11 years	3	18	3	18
12 - 15 years	2	12	2	12
16 years or more	1	6	1	6
I am not a school leader	8	47	7	41
<b>If you are building level school leader, what grade levels are taught in your school?</b>				
Elementary	5	29	3	18
Middle School	2	12	0	0
High School	2	12	5	29
I am not a school leader	8	47	9	53

**Table D1 (continued)*****Panel Member Demographics (by Panel)***

	<b>Panel 1</b>		<b>Panel 2</b>	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
<b>If you are building -level school leader, which best describes the location of your school?</b>				
Urban	1	6	3	18
Suburban	3	18	2	12
Rural	5	29	3	18
I am not a school leader	8	47	9	53
<b>Are you currently involved in the training or preparation of school leaders?</b>				
Yes	8	47	7	41
No	0	0	0	0
I am not college faculty	9	53	10	59
<b>How many years of experience (including this year) do you have preparing school leaders?</b>				
3 years or less	0	0	0	0
4 - 7 years	0	0	0	0
8 - 11 years	3	18	1	6
12 - 15 years	2	12	1	6
16 years or more	3	18	5	29
Not college faculty	9	53	10	59

**Table D2**  
*Passing Score Summary by Round of Judgments*

<b>Panelist</b>	<b>Panel 1</b>		<b>Panel 2</b>	
	<b>Round 1</b>	<b>Round 2</b>	<b>Round 1</b>	<b>Round 2</b>
1	69.51	71.69	86.32	85.72
2	73.24	73.24	72.36	70.87
3	66.27	66.27	65.77	65.47
4	84.86	83.46	72.11	73.39
5	64.63	68.96	70.34	70.54
6	87.21	84.11	66.13	74.33
7	80.77	82.37	68.47	71.56
8	87.90	87.80	92.28	83.79
9	72.53	75.22	79.99	79.64
10	68.74	70.94	69.74	70.14
11	68.26	72.03	90.62	90.32
12	74.84	75.23	67.37	71.73
13	76.73	77.73	72.83	72.83
14	81.32	79.63	82.91	83.51
15	85.91	85.01	77.99	78.99
16	75.66	73.56	75.01	75.11
17	74.54	74.64	83.93	83.83
<b>Average</b>	76.06	76.58	76.13	76.58
<b>Lowest</b>	64.63	66.27	65.77	65.47
<b>Highest</b>	87.90	87.80	92.28	90.32
<b>SD</b>	7.49	6.19	8.51	6.87
<b>SEJ</b>	1.82	1.50	2.06	1.67

**Table D3*****Final Evaluation: Panel 1***

	<b>Strongly agree</b>		<b>Agree</b>		<b>Disagree</b>		<b>Strongly disagree</b>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• I understood the purpose of this study.	16	94	1	6	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	13	76	4	24	0	0	0	0
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	15	88	2	12	0	0	0	0
• The explanation of how the recommended cut score is computed was clear.	14	82	3	18	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	15	88	2	12	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	14	82	3	18	0	0	0	0
• I understood how to use the survey software.	14	82	3	18	0	0	0	0

**Table D3 (continued)**  
**Final Evaluation: Panel 1**

<b>How influential was each of the following factors in guiding your standard-setting judgments?</b>	<b>Very influential</b>		<b>Somewhat influential</b>		<b>Not influential</b>			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• The description of the just qualified candidate	12	71	5	29	0	0		
• The between-round discussions	11	65	6	35	0	0		
• The knowledge/skills required to answer each test question	14	82	3	18	0	0		
• The cut scores of other panel members	5	29	12	71	0	0		
• My own professional experience	9	53	8	47	0	0		
	<b>Very comfortable</b>		<b>Somewhat comfortable</b>		<b>Somewhat uncomfortable</b>		<b>Very uncomfortable</b>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• Overall, how comfortable are you with the panel's recommended cut score?	13	76	3	18	1	6	0	0
	<b>Too low</b>		<b>About right</b>		<b>Too high</b>			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended cut score is:	2	12	15	88	0	0		

**Table D4*****Final Evaluation: Panel 2***

	<b>Strongly agree</b>		<b>Agree</b>		<b>Disagree</b>		<b>Strongly disagree</b>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• I understood the purpose of this study.	17	100	0	0	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	14	82	3	18	0	0	0	0
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	14	82	3	18	0	0	0	0
• The explanation of how the recommended cut score is computed was clear.	12	71	5	29	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	15	88	2	12	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	13	76	4	24	0	0	0	0
• I understood how to use the survey software.	16	94	1	6	0	0	0	0

**Table D4 (continued)**  
**Final Evaluation: Panel 2**

<b>How influential was each of the following factors in guiding your standard-setting judgments?</b>	<b>Very influential</b>		<b>Somewhat influential</b>		<b>Not influential</b>			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• The description of the just qualified candidate	12	71	5	29	0	0		
• The between-round discussions	8	47	9	53	0	0		
• The knowledge/skills required to answer each test question	14	82	3	18	0	0		
• The cut scores of other panel members	6	35	11	65	0	0		
• My own professional experience	14	82	3	18	0	0		
	<b>Very comfortable</b>		<b>Somewhat comfortable</b>		<b>Somewhat uncomfortable</b>		<b>Very uncomfortable</b>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• Overall, how comfortable are you with the panel's recommended cut score?	14	82	3	18	0	0	0	0
	<b>Too low</b>		<b>About right</b>		<b>Too high</b>			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended cut score is:	0	0	17	100	0	0		