



*Listening. Learning. Leading.*

Multistate Standard-Setting Technical Report

**PRAXIS™ MIDDLE SCHOOL SCIENCE (5440)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2014

---

# EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ Middle School Science (5440) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

## PARTICIPATING STATES

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

## 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 Praxis Middle School Science test, the recommended passing score<sup>1</sup> is 60 out of a possible 100 raw-score points. The scaled score associated with a raw score of 60 is 150 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 Praxis™ Middle School Science (5440) test, research staff from ETS designed and conducted a multistate standard-setting study in February 2014 in Princeton, New Jersey. Education agencies<sup>2</sup> recommended panelists with (a) experience as either science teachers or college faculty who prepare science teachers and (b) familiarity with the knowledge and skills required of beginning science teachers. Twenty states and Guam (Table 1) were represented by 32 panelists. (See Appendix A for the names and affiliations of the panelists.)

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

---

Arkansas (2 panelists)	Nevada (1 panelist)
Delaware (1 panelist)	New Hampshire (1 panelist)
Guam (1 panelist)	New Jersey (2 panelists)
Hawaii (1 panelist)	North Carolina (1panelist)
Idaho (2 panelists)	North Dakota (2 panelists)
Kansas (1 panelist)	Rhode Island (1 panelist)
Kentucky (2 panelists)	South Carolina (2 panelists)
Louisiana (2 panelists)	South Dakota (2 panelists)
Maine (1 panelist)	Virginia (2 panelists)
Maryland (2 panelists)	West Virginia (1 panelist)
Mississippi (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 represents the combined judgments of two panels of experienced educators. Each jurisdiction may want

---

<sup>2</sup> States and jurisdictions that currently use Praxis 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.

to consider the recommended passing score but also other sources of information when setting the final Praxis Middle School Science 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 Praxis Middle School Science test 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 Praxis Middle School Science test 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 PRAXIS MIDDLE SCHOOL SCIENCE TEST

The Praxis Middle School Science *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level science teachers have the knowledge/skills believed necessary for competent professional practice.

The two and a half-hour test contains 125 selected-response<sup>4</sup> items covering six content areas: *Scientific Inquiry, Methodology, Techniques, and History* (approximately 15 items), *Basic Principles of Matter and Energy* (approximately 15 items), *Physical Sciences* (approximately 28 items), *Life Sciences* (approximately 30 items), *Earth and Space Sciences* (approximately 22 items), and *Science, Technology, and Society* (approximately 15 items).<sup>5</sup> The reporting scale for the Praxis Middle School Science ranges from 100 to 200 scaled-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.

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.

### REVIEWING THE TEST

The standard-setting panelists first reviewed 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.

---

<sup>4</sup> Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate's score.

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

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 teachers or areas that address content particularly important for entry-level teachers.

## DESCRIBING 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 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 standard-setting process for the Praxis Middle School Science 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 and decided if, overall, the item would be

difficult for the just qualified candidate, easy for the just qualified candidate or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Difficult items for the just qualified candidate are in the 0 to .30 range.
- Moderately difficult/easy items for the just qualified candidate are in the .40 to .60 range.
- Easy items for the just qualified candidate are in the .70 to 1 range.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that an item would be easy for the just qualified candidate, the initial decision located the item in the .70 to 1 range. The second decision for the panelist was to decide 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.

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. 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.

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, including the summary of the Round 1 judgments, 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 32 educators representing 20 states and Guam. (See Appendix A for a listing of panelists.) Seventeen panelists were teachers, eleven were college faculty, one was an administrator or department head, and three held another position. All of the faculty members’ job responsibilities included the training of science teachers.

Table D1 (in Appendix D) presents a summary of demographic information by panel.

**Table 2**

*Panel Member Demographics (Across Panels)*

	<i>N</i>	<i>%</i>
<b>Current position</b>		
Teacher	17	53%
Administrator/Department head	1	3%
College faculty	11	34%
Other	3	9%
<b>Race</b>		
White	20	63%
Black or African American	5	16%
Hispanic or Latino	2	6%
Asian or Asian American	2	6%
American Indian or Alaskan Native	1	3%
Native Hawaiian or Other Pacific Islander	1	3%
Other	1	3%

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

	<i>N</i>	<i>%</i>
<b>Gender</b>		
Female	23	72%
Male	9	28%
<b>Are you currently certified to teach this subject in your state?</b>		
Yes	27	84%
No	5	16%
<b>Are you currently teaching this subject in your state?</b>		
Yes	25	78%
No	7	22%
<b>Are you currently supervising or mentoring other teachers of this subject?</b>		
Yes	18	56%
No	14	44%
<b>At what K–12 grade level are you currently teaching this subject?</b>		
Middle School (6 - 8 or 7 - 9)	17	53%
High School (9 – 12 or 10 - 12)	1	3%
Middle and High School	1	3%
Not currently teaching at the K–12 level	13	41%
<b>Including this year, how many years of experience do you have teaching this subject?</b>		
3 years or less	2	6%
4–7 years	4	13%
8–11 years	11	34%
12–15 years	4	13%
16 years or more	11	34%
<b>Which best describes the location of your K–12 school?</b>		
Urban	6	19%
Suburban	7	22%
Rural	6	19%
Not currently working at the K–12 level	13	41%
<b>If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?</b>		
Yes	11	34%
No	0	0%
Not college faculty	21	66%

## 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>6</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	62.09	57.86
Lowest	51.60	49.20
Highest	71.85	69.90
SD	5.21	5.56
SEJ	1.30	1.39

---

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.

---

<sup>6</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).

The panels' passing score recommendations for the Praxis Middle School Science are 62.09 for Panel 1 and 57.86 for Panel 2 (out of a possible 100 raw-score points). The values were rounded to the next highest whole number, to determine the functional recommended passing score — 63 for Panel 1 and 58 for Panel 2. The scaled scores associated with 63 and 58 raw points are 154 and 147, respectively.

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 Praxis Middle School Science is 59.98 (out of a possible 100 raw-score points). The value was rounded to 60 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 60 raw points is 150.

Table 4 presents the estimated conditional standard error of measurement (CSEM) around the recommended passing score. A standard error represents the uncertainty associated with a test score. The scaled scores associated with one and two CSEMs 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 CSEMs of the Recommended Passing Score<sup>7</sup>*

<b>Recommended passing score (CSEM)</b>		<b>Scale score equivalent</b>
	60 (4.92)	150
-2 CSEMs	51	137
-1 CSEM	56	144
+ 1 CSEM	65	157
+ 2 CSEMs	70	164

**Note.** CSEM = conditional standard error of measurement.

<sup>7</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 scaled scores.

## 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.

All panelists *strongly agreed* or *agreed* that they understood the purpose of the study and 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 but two of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 23 of the 32 panelists were *very comfortable*. Twenty-nine of the 32 panelists indicated the recommended passing score was *about right* with the three remaining panelists indicating that the score was *too low*.

---

## SUMMARY

To support the decision-making process for education agencies establishing a passing score (cut score) for the Praxis Middle School Science, 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 Praxis Middle School Science, the recommended passing score<sup>8</sup> is 60 out of a possible 100 raw-score points. The scaled score associated with a raw score of 60 is 150 on a 100–200 scale.

---

<sup>8</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.
- ETS. (In Press). *The Praxis Series: Middle School Science (5440)*. 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.
- 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.

---

## APPENDIX A

# PANELISTS' NAMES & AFFILIATIONS

### *Participating Panelists With Affiliations*

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Nancy Allen	Gilford Middle School (NH)
Katie Anderson	East Middle School-Rapid City School District (SD)
Gena Asevado	N. P. Trist Middle School (LA)
Steve Beckelhimer	Marshall University/June Harless Center (WV)
April Bullen	McCullough Middle School (DE)
Laurie Cleavinger	University of Kansas (KS)
Michelle Crane	Smee School District (SD)
André E. DeLeón	Nevada Dept. of Education (NV)
Kyle Engdahl	Cheney Middle School (ND)
Janice Francis	PCSSD, Maumelle Middle School (AR)
Esther Frazier	Madison Middle School (MS)
Chelsey Gravseth	Mandan Middle School (ND)
Albert Hayward	South Carolina State University (SC)
Allen Henderson	Harding University (AR)
Ebony Hill	Richmond Public Schools-Thompson Middle School (VA)
Lisa Hopkins	Dorchester County Public Schools (MD)
Richard Jones	University of Hawaii West Oahu (HI)
Misti Kelly	Stevenson University (MD)
John Labriola	Chariho Middle School (RI)
Carole Lee	University of Maine at Farmington (ME)
Rachel Lowery	Kings Mountain Intermediate School (NC)
Michiko McClary	Clafin University (SC)
Dennis McDill	Berwick High School (LA)
Renu Mendiratta	John Adams Middle School, Edison (NJ)
Elizabeth Morales	New Brunswick Public Schools (NJ)
Louis Nadelson	Boise State University (ID)
Kimberly Riggs-Poole	Hampton City Schools (VA)
Wendy Ruchti	Idaho State University (ID)
Cheryl Sanguenza	University of Guam (GU)
Melissa L. Shirley	University of Louisville (KY)
Sandra Thomas-Jenkins	Madison Middle School (MS)
Corean Wells	Corbin Middle School (KY)

---

---

APPENDIX B  
STUDY AGENDA

# **AGENDA**

## **Praxis Middle School Science (5440) Standard-Setting Study**

Day 1

Welcome and Introduction

Overview of Standard Setting and the Praxis Middle School Science test

Review the test

Discuss test

Break

Describe the Knowledge/Skills of a Just Qualified Candidate

Create the JQC description

Lunch

Create the JQC description (continued)

Break

Discuss and finalize JQC description

Training for standard-setting judgments

Practice judgments & discuss

Round 1 judgments

Collect Materials; End of Day 1

# **AGENDA**

## **Praxis Middle School Science (5440) Standard-Setting Study**

Day 2

Review Day 1 & Preview Day 2

Round 1 standard setting judgments (continued from Day 1)

Discuss judgments & Round 2

Lunch

Discuss judgments & Round 2 (continued)

Complete final evaluation

Collect materials; End of study

---

## APPENDIX C

### JUST QUALIFIED CANDIDATE DESCRIPTION

## Description of the Just Qualified Candidate<sup>9</sup>

### A Just Qualified Candidate ...

#### I. Scientific Inquiry, Methodology, Techniques, and History

- A. Understands methods of scientific inquiry and how methods are used in basic problem solving
- B. Understands the processes involved in scientific data collection and manipulation
- C. Understands how to interpret and draw conclusions from data presented in tables, graphs, and charts

#### II. Basic Principles of Matter and Energy

- A. Knows the structure and properties of matter; can identify occurrence and abundance of the elements and their isotopes
- B. Knows the basic relationships between energy and matter and identifies the basic relationships between energy and matter
- C. Knows the basic structure of the atom; can identify ions, electron arrangements, radioactivity, and applications of radioactivity

#### III. Physical Sciences

##### A. Physics

- 1. Knows mechanics and knows linear motion in 1 & 2 dimensions (speed, velocity, acceleration), including—distinguish between mass and weight, conservation of energy, and density
- 2. Can identify circular motion in 1 & 2 dimensions, inertia vs. momentum, simple machines and mechanical advantage, buoyancy and pressure
- 3. Can identify electricity & magnetism
- 4. Can identify basic waves and optics

##### B. Chemistry

- 1. Knows organization of periodic table and how to use it to predict physical & chemical properties
- 2. Can identify covalent and ionic bonding
- 3. Knows names of simple chemical compounds
- 4. Knows how to balance simple chemical equations
- 5. Can identify factors that affect reaction rates
- 6. Can identify chemical and physical properties of acids and bases
- 7. Can identify pH scale and neutralization
- 8. Can identify dilute vs. concentrated solutions
- 9. Can identify difference between solute and solvent
- 10. Can identify effect of temperature and particle size on dissolving

---

<sup>9</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>10</sup> (continued)

### A Just Qualified Candidate ...

#### IV. Life Sciences

1. Understands the basic structure, function of cells, cellular organelles and cell reproduction
2. Knows basic genetics including DNA structure, replication, Mendalian inheritance, mutations and can identify common genetic disorders
3. Knows the major structures and functions of plant and animals organs and systems
4. Knows key aspects of ecology
5. Knows the theory and key mechanisms of evolution

#### V. Earth and Space Sciences

1. Understands the basics of the rock cycle and knows plate tectonics (e.g., weathering, erosion, deposition, earthquakes, and volcanoes)
2. Knows properties of water and identifies structures and processes of earth's oceans and bodies of water
3. Knows basic meteorology (e.g., Earth's atmosphere, frontal systems and precipitation) and can identify major factors that affect climate and seasons
4. Understands interaction of Earth, moon, sun and identifies major features of solar system and universe

#### VI. Science, Technology, and Society

1. Understands the impact of science and technology and major issues on the environment and management of natural resources
2. Can identify applications of science and technology in daily life and the impact of science on public-health issues

---

<sup>10</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>				
Teacher	9	56%	8	50%
Administrator/Department head	0	0%	1	6%
College faculty	5	31%	6	38%
Other	2	13%	2	6%
<b>Race</b>				
White	10	63%	10	63%
Black or African American	3	19%	2	13%
Hispanic or Latino	1	6%	1	6%
Asian or Asian American	1	6%	1	6%
American Indian or Alaskan Native	0	0%	1	6%
Native Hawaiian or Other Pacific Islander	0	0%	1	6%
Other	1	6%	0	0%
<b>Gender</b>				
Female	11	69%	12	75%
Male	5	31%	4	25%
<b>Are you currently certified to teach this subject in your state?</b>				
Yes	14	88%	13	81%
No	2	13%	3	19%
<b>Are you currently teaching this subject in your state?</b>				
Yes	13	81%	12	75%
No	3	19%	4	25%
<b>Are you currently supervising or mentoring other teachers of this subject?</b>				
Yes	6	38%	12	75%
No	10	63%	4	25%
<b>At what K–12 grade level are you currently teaching this subject?</b>				
Middle School (6–8 or 7–9)	9	56%	8	50%
High School (9–12 or 10–12)	1	6%	0	0%
Middle and High School	1	6%	0	0%
Not currently teaching at the K–12 level	5	31%	8	50%

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

	<b>Panel 1</b>		<b>Panel 2</b>	
	<i>N</i>	%	<i>N</i>	%
<b>Including this year, how many years of experience do you have teaching this subject?</b>				
3 years or less	1	6%	1	6%
4–7 years	4	25%	0	0%
8–11 years	5	31%	6	38%
12–15 years	1	6%	3	19%
16 years or more	5	31%	6	38%
<b>Which best describes the location of your K–12 school?</b>				
Urban	3	19%	3	19%
Suburban	3	19%	4	25%
Rural	4	25%	2	13%
Not currently working at the K–12 level	6	38%	7	44%
<b>If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?</b>				
Yes	5	31%	6	38%
No	0	0%	0	0%
Not college faculty	11	69%	10	63%

**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	50.50	51.60	47.60	49.20
2	64.90	65.00	65.90	62.55
3	71.60	71.85	53.70	52.90
4	63.25	62.65	56.60	56.30
5	64.75	65.25	63.20	63.55
6	60.55	62.95	64.90	64.90
7	54.95	57.70	48.75	50.35
8	57.35	58.05	74.10	69.90
9	65.80	65.25	55.00	55.30
10	60.60	62.00	54.10	54.50
11	63.75	62.65	57.15	58.25
12	56.40	57.90	62.10	61.80
13	64.50	64.90	58.00	56.70
14	56.55	58.50	55.85	57.45
15	54.35	56.80	60.00	58.70
16	72.25	70.45	48.95	53.35
<b>Average</b>	61.38	62.09	57.87	57.86
<b>Lowest</b>	50.50	51.60	47.60	49.20
<b>Highest</b>	72.25	71.85	74.10	69.90
<b>SD</b>	6.11	5.21	7.04	5.56
<b>SEJ</b>	1.53	1.30	1.76	1.39

**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.	15	94%	1	6%	0	0%	0	0%
• The instructions and explanations provided by the facilitator were clear.	12	75%	4	25%	0	0%	0	0%
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	13	81%	3	19%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	6	38%	7	44%	3	19%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	11	69%	5	31%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	11	69%	5	31%	0	0%	0	0%

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

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential		Somewhat influential		Not influential			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• The description of the just qualified candidate	16	100%	0	0%	0	0%		
• The between-round discussions	8	50%	8	50%	0	0%		
• The knowledge/skills required to answer each test item	11	69%	5	31%	0	0%		
• The passing scores of other panel members	1	6%	9	56%	6	38%		
• My own professional experience	7	44%	8	50%	1	6%		
	Very comfortable		Somewhat comfortable		Somewhat uncomfortable		Very uncomfortable	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• Overall, how comfortable are you with the panel's recommended passing score?	10	63%	5	31%	1	6%	0	0%
	Too low		About right		Too high			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended passing score is:	2	13%	14	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.	16	100%	0	0%	0	0%	0	0%
• The instructions and explanations provided by the facilitator were clear.	15	94%	1	6%	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	94%	1	6%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	11	69%	5	31%	0	0%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	14	88%	2	13%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	13	81%	3	19%	0	0%	0	0%

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

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential		Somewhat influential		Not influential			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• The description of the just qualified candidate	13	81%	3	19%	0	0%		
• The between-round discussions	8	50%	7	44%	1	6%		
• The knowledge/skills required to answer each test item	13	81%	3	19%	0	0%		
• The passing scores of other panel members	3	19%	11	69%	2	13%		
• My own professional experience	13	81%	3	19%	0	0%		
	Very comfortable		Somewhat comfortable		Somewhat uncomfortable		Very uncomfortable	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• Overall, how comfortable are you with the panel's recommended passing score?	13	81%	2	13%	1	6%	0	0%
	Too low		About right		Too high			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended passing score is:	1	6%	15	94%	0	0%		