Measuring the Power of Learning.™

Multistate Standard-Setting Technical Report

PRAXIS® MIDDLE SCHOOL SCIENCE (5442)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

January 2020

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Middle School Science (5442) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING STATES

Panelists from 17 states and Washington, D.C. were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either middle school science teachers or college faculty who prepare middle school science teachers and (b) familiarity with the knowledge and skills required of beginning middle school 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[1] is 61 out of a possible 100 raw-score points. The scale score associated with a raw score of 61 is 152 on a 100–200 scale.

[1] 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 (5442) test, research staff from ETS designed and conducted a multistate standard-setting study in December 2019 in Princeton, New Jersey. Education agencies[2] recommended panelists with (a) experience as either middle school science teachers or college faculty who prepare middle school science teachers and (b) familiarity with the knowledge and skills required of beginning middle school science teachers. Seventeen states and Washington, D.C. (Table 1) were represented by 31 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1

Participating States and Washington, D.C. and Number of Panelists

Arkansas (1 panelist)	New Jersey (2 panelists)
Hawaii (1 panelist)	New Mexico (1 panelist)
Idaho (2 panelists)	Rhode Island (2 panelists)
Indiana (3 panelists)	South Carolina (1 panelist)
Kentucky (2 panelists)	South Dakota (2 panelists)
Louisiana (1 panelist)	Utah (2 panelists)
Maryland (1 panelist)	Virginia (2 panelists)
North Carolina (2 panelists)	Washington, D.C. (2 panelists)
Nebraska (2 panelists)	West Virginia (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 state and D.C., 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,[3] which represents the combined judgments of two panels of experienced educators. Each state, or D.C., may want 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 state, or D.C., may accept the recommended passing score, adjust the score upward to reflect more stringent

[2] States and jurisdictions that currently use *Praxis* tests were invited to participate in the multistate standard-setting study.

[3] In addition to the recommended passing score averaged across the two panels, the recommended passing scores for each panel are presented.

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 state's, or D.C.'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 states, and D.C., 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 states, and D.C., 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 state, and D.C., 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 possesses the required knowledge/skills. States, and D.C., need to consider which decision error is more important to minimize.

OVERVIEW OF THE *PRAXIS*® MIDDLE SCHOOL SCIENCE TEST

The Praxis® Middle School Science *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures the knowledge and competencies necessary for a beginning teacher of middle school science.

The two and a half-hour assessment contains 125 multiple-choice items[4] covering four content areas: *Nature and Impact of Science and Engineering* (approximately 17 items), *Physical Science* (approximately 38 items), *Life Science* (approximately 38 items), *and Earth and Space Science* (approximately 32 items).[5] The reporting scale for the *Praxis* Middle School Science test ranges from 100 to 200 scale-score points.

PROCESSES AND METHODS

The design of the standard-setting study included two, independent 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 as a general session for both panels. The session opened with a welcome and introduction by each of the meeting facilitators. The facilitators described the test, provided an overview of standard setting, and presented the agenda for the study. Appendix B shows the standard-setting study agenda.

REVIEWING THE TEST

While both panels were together during the general session, the standard-setting panelists took the test and then discussed the content measured. 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.

[4] Twenty-five of the 125 multiple-choice items are pretest items and do not contribute to a candidate's score.
[5] 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.

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.

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 standard-setting process for the *Praxis* Middle School Science test 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.

For consistency in understanding the standard-setting judgment process, both panels remained together as they received training and practice in how to complete their standard-setting judgments. 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 determined what was the probability that the

just qualified candidate would answer 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 rationales. All panelists completed a post-training evaluation 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. After the independent judgments were completed, panelists were split into two, distinct panels that worked separately for the remainder of the study.

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 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 31 educators representing 17 states and D.C. (See Appendix A for a listing of panelists.) Eighteen panelists were teachers, eight were college faculty, one was an administrator or department head, and four held another position. All eight of the faculty members' job responsibilities included the training of middle school science teachers.

The number of experts by panel and their demographic information are presented in Appendix D (Table D1).

Table 2

Panel Member Demographics (Across Panels)

	N	%
Current position		
Teacher	18	58
Administrator/Department head	1	3
College faculty	8	26
Other	4	13
Race		
White or European American	22	71
Black or African American	4	13
Hispanic or Latino	3	10
Asian or Asian American	1	3
African American/Caucasian	1	3
Gender		
Female	19	61
Male	12	39
Are you currently certified as a teacher of this subject in your state?		
Yes	27	87
No	4	13
Are you currently teaching this subject in your state?		
Yes	23	74
No	8	26
Are you currently supervising or mentoring other teachers of this subject?		
Yes	24	77
No	7	23
At what K–12 grade level are you currently teaching this subject?		
Middle school (6–8 or 7–9)	12	39
High school (9–12 or 10–12)	4	13
Middle and High school	2	6
Not currently teaching at the K–12 level	13	42

Table 2 (continued)

Panel Member Demographics (Across Panels)

	N	%
Including this year, how many years of experience do you have teaching this subject?		
3 years or less	2	6
4–7 years	2	6
8–11 years	11	35
12–15 years	5	16
16 years or more	11	35
Which best describes the location of your K–12 school?		
Urban	5	16
Suburban	4	13
Rural	9	29
Not currently working at the K–12 level	13	42
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?		
Yes	8	26
No	0	0
Not college faculty	23	74

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.[6] 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).

[6] 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).

Table 3
Summary of Round 2 Standard-setting Judgments

	Panel 1	Panel 2
Average	60.58	60.33
Lowest	48.00	55.65
Highest	73.20	67.25
SD	6.84	3.25
SEJ	1.71	0.84

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 *Praxis* Middle School Science test are 60.58 for Panel 1 and 60.33 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 — 61 for both Panel 1 and Panel 2. The scale score associated with 61 raw points is 152.

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 test is 60.46 (out of a possible 100 raw-score points). The value was rounded to 61 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 61raw points is 152.

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 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[7]

Recommended passing score (CSEM)		Scale score equivalent
61 (4.90)		152
- 2 CSEM	52	140
- 1 CSEM	57	147
+1 CSEM	66	159
+2 CSEM	71	166

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

FINAL EVALUATIONS

The panelists completed an evaluation at the conclusion of the 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 the 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. Thirty of the 31 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; 20 of the 31 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. More than half of the panelists (25 of the 31 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

All of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 28 of the 31 panelists were *very comfortable*. Twenty-nine of the 31 panelists

[7] 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.

indicated the recommended passing score was *about right*; one panelist indicated that the passing score was *too low*, and one panelist indicated that the passing score was *too high*.

SUMMARY

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Middle School Science test, 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 test, the recommended passing score[8] is 61 out of a possible 100 raw-score points. The scale score associated with a raw score of 61 is 152 on a 100–200 scale.

[8] 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®: The Praxis Study Companion: Middle School Science (5442).* 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 Affiliation

Panelist	**Affiliation**
Mollie Craven	Elkins High School (WV)
Teresa Dobler	Washington Latin PCS (DC)
Shannon Furstenau	Lincoln Public School (NE)
Ehren Haderlie	Brigham Young University-Idaho (ID)
Breanna Hagedorn	Phillip School Middle School (Lincoln Public Schools) (NE)
Cecilia Hernandez	New Mexico State University (NM)
Shawn Hicks	Harnett County Schools/Harnett Central Middle School (NC)
Ellie Houghton	John Griffin Middle School (NC)
Belinda Jenkins	Edward W. Wyatt Middle School (VA)
Alicia Killean	Holmdel Township Public Schools (NJ)
Chantel Kornegay	District of Columbia Public Schools (DC)
John Labriola	Chariho Middle School (RI)
Daniel Levin	University of Maryland, College Park (MD)
Alvin Lin	Central District, Hawaii Department of Education (HI)
Angela Marksberry	Henderson North Middle School (KY)
Lloyd Mataka	Lewis-Clark State College (ID)
Pamela Medows	Taylor University (IN)
Lisa Nance	Caddo Parish Schools (LA)
Matthew Perkins Coppola	Purdue University-Fort Wayne (IN)
Chad Ronish	Hill City High School (SD)
Adam Scott	Archie R. Cole Middle School (RI)

Participating Panelists With Affiliation (continued)

Panelist	**Affiliation**
Gwendolynn Shealy	South Carolina Department of Education (SC)
Angela Stanford	Southern Arkansas University (AR)
Angela Stewart	Centennial Junior High, Davis School District (UT)
Jamaal Stiles	Marion County Public Schools (KY)
Melesa Swartz	West Liberty University (WV)
Nora Walsh	FJ Reitz High School (IN)
Pam Wells	Mobridge-Pollock School District (SD)
Cheryl Zanone	Robert R. Lazar Middle School (NJ)
Bernard Zdancewicz	Greensville County Public Schools (VA)

*One panelist did not wish to be listed in the final report.

APPENDIX B

STUDY AGENDA

AGENDA

Praxis® Middle School Science (5442) Standard-Setting Study

Day 1

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Middle School Science Test

Review the *Praxis* Middle School Science Test

Discuss the *Praxis* Middle School Science Test

Break

Define the Knowledge/Skills of a Just Qualified Candidate

Lunch

Define the Knowledge/Skills of a Just Qualified Candidate (continued)

Break

Standard-Setting Training

Round 1 Standard Setting Judgments

Collect Materials; End of Day 1

AGENDA

Praxis® Middle School Science (5442)
Standard-Setting Study

Day 2

Overview of Day 2

Round 1 Standard Setting Judgments (continued)

Break

Round 1 Feedback and Round 2 Judgments

Lunch

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[9]

A just qualified candidate …

I. Nature and Impact of Science and Engineering
   1. Understands that science uses evidence gathered through a variety of methods as a foundation of scientific knowledge
   2. Is familiar with the similarities and difference between science and engineering practices
   3. Recognizes common safety concerns and practices for general lab procedures, equipment use, and material storage
   4. Knows science, engineering, and technology drive each other forward and impact the environment and society

II. Physical Science
   1. Knows energy can be transferred within and between systems and is transformed from one form to another
   2. Knows that differences in structure determine properties of matter and that those properties cause differences in chemical and physical properties and changes
   3. Knows basic relationships between forces and motion
   4. Knows the difference between matter and energy

III. Life Science
   1. Knows the structures and processes associated with cellular function and specialization (i.e., tissues, organs, systems)
   2. Knows the interactions between the abiotic and biotic components within the environment
   3. Knows the effects of genetic and environmental factors on the diversity of life

IV. Earth and Space Science
   1. Knows predictable patterns and relationships in the Sun-Earth-Moon and effect on Earth
   2. Knows the processes of the formation and changes in rocks (e.g., due to plate tectonics, weathering, and erosion)
   3. Knows how atmospheric patterns impact weather
   4. Knows causes of climate change (natural and human)

[9] 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)

	Panel 1		Panel 2	
	N	**%**	***N***	**%**
Current position				
Teacher	9	56	9	60
Administrator/Department head	0	0	1	7
College faculty	5	31	3	20
Other	2	13	2	13
Race				
White or European American	11	69	11	73
Black or African American	2	13	2	13
Hispanic or Latino	2	13	1	7
Asian or Asian American	0	0	1	7
African American/Caucasian	1	6	0	0
Gender				
Female	11	69	8	53
Male	5	31	7	47
Are you currently certified as a teacher of this subject in your state?				
Yes	14	88	13	87
No	2	13	2	13
Are you currently teaching this subject in your state?				
Yes	11	69	12	80
No	5	31	3	20
Are you currently supervising or mentoring other teachers of this subject?				
Yes	12	75	12	80
No	4	25	3	20
At what K–12 grade level are you currently teaching this subject?				
Middle school (6–8 or 7–9)	6	38	6	40
High school (9–12 or 10–12)	2	13	2	13
Middle and High school	1	6	1	7
Not currently teaching at the K–12 level	7	44	6	40

Table D1 (continued)

Panel Member Demographics (by Panel)

	Panel 1		Panel 2	
	N	**%**	***N***	**%**
Including this year, how many years of experience do you have teaching this subject?				
3 years or less	2	13	0	0
4–7 years	0	0	2	13
8–11 years	6	38	5	33
12–15 years	3	19	2	13
16 years or more	5	31	6	40
Which best describes the location of your K–12 school?				
Urban	3	19	2	13
Suburban	2	13	2	13
Rural	4	25	5	33
Not currently working at the K–12 level	7	44	6	40
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?				
Yes	5	31	3	20
No	0	0	0	0
Not college faculty	11	69	12	80

Table D2

Passing Score Summary by Round of Judgments

	Panel 1		Panel 2	
Panelist	**Round 1**	**Round 2**	**Round 1**	**Round 2**
1	53.55	54.15	61.90	62.60
2	72.85	73.20	60.45	61.10
3	63.35	61.95	56.10	56.80
4	63.90	64.55	55.20	56.50
5	63.35	62.45	70.90	67.25
6	60.30	62.05	61.90	60.70
7	66.05	65.55	57.05	58.95
8	73.55	71.15	59.75	62.25
9	65.00	63.80	65.30	65.65
10	45.30	48.00	58.85	59.30
11	56.70	55.00	58.30	60.50
12	54.80	55.60	60.30	61.10
13	46.00	55.35	57.55	57.55
14	52.90	51.70	55.05	55.65
15	59.25	60.75	58.70	59.10
16	64.40	63.95		
Average	60.08	60.58	59.82	60.33
Lowest	45.30	48.00	55.05	55.65
Highest	73.55	73.20	70.90	67.25
SD	8.21	6.84	4.12	3.25
SEJ	2.05	1.71	1.06	0.84

Table D3

Final Evaluation: Panel 1

	Strongly agree		Agree		Disagree		Strongly disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	13	81	3	19	0	0	0	0
• The instructions and explanations provided by the facilitators 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.	14	88	2	13	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	15	94	1	6	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	15	94	1	6	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	12	75	4	25	0	0	0	0
• I understood how to use the survey software.	15	94	1	6	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	
	N	%	*N*	%	*N*	%
• The description of the just qualified candidate	12	75	4	25	0	0
• The between-round discussions	12	75	4	25	0	0
• The knowledge/skills required to answer each test item	14	88	2	13	0	0
• The passing scores of other panel members	8	50	7	44	1	6
• My own professional experience	14	88	2	13	0	0

	Very comfortable		Somewhat comfortable		Somewhat uncomfortable		Very uncomfortable	
	N	%	*N*	%	*N*	%	*N*	%
• Overall, how comfortable are you with the panel's recommended passing score?	13	81	3	19	0	0	0	0

	Too low		About right		Too high	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	1	6	14	88	1	6

Table D4

Final Evaluation: Panel 2

	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 facilitators were clear.	11	73	3	20	1	7	0	0
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	12	80	3	20	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	13	87	2	13	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	15	100	0	0	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	12	80	3	20	0	0	0	0
• I understood how to use the survey software.	15	100	0	0	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	
	N	%	*N*	%	*N*	%
• The description of the just qualified candidate	8	53	7	47	0	0
• The between-round discussions	12	80	3	20	0	0
• The knowledge/skills required to answer each test item	13	87	2	13	0	0
• The passing scores of other panel members	5	33	10	67	0	0
• My own professional experience	11	73	4	27	0	0

	Very comfortable		Somewhat comfortable		Somewhat uncomfortable		Very uncomfortable	
	N	%	*N*	%	*N*	%	*N*	%
• Overall, how comfortable are you with the panel's recommended passing score?	15	100	0	0	0	0	0	0

	Too low		About right		Too high	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0	15	100	0	0