



Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

PRAXIS™ EARLY CHILDHOOD EDUCATION (5025)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2015

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ Early Childhood Education (5025) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING STATES

Panelists from 16 states and Washington, D.C. were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either early childhood teachers or college faculty who prepare early childhood teachers and (b) familiarity with the knowledge and skills required of beginning early childhood 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 Early Childhood Education test, the recommended passing score¹ is 64 out of a possible 100 raw-score points. The scaled score associated with a raw score of 64 is 156 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™ Early Childhood Education (5025) test, research staff from ETS designed and conducted a multistate standard-setting study in February 2015 in Princeton, New Jersey. Education agencies² recommended panelists with (a) experience as either early childhood teachers or college faculty who prepare early childhood teachers and (b) familiarity with the knowledge and skills required of beginning early childhood teachers. Sixteen states and Washington, D.C. (Table 1) were represented by 38 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating Jurisdictions and Number of Panelists

Alabama (3 panelists)	North Dakota (1 panelist)
Arkansas (3 panelists)	New Hampshire (2 panelists)
Hawaii (2 panelists)	New Jersey (4 panelists)
Idaho (1 panelist)	Nevada (2 panelists)
Iowa (2 panelists)	Rhode Island (1 panelist)
Kentucky (4 panelists)	Utah (2 panelists)
Maryland (2 panelists)	Washington, D.C. (2 panelists)
Montana (2 panelists)	Wyoming (1 panelist)
North Carolina (4 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,³ which 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

² States and jurisdictions that currently use Praxis 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.

Praxis Early Childhood Education 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 Early Childhood Education 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 Early Childhood Education 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 EARLY CHILDHOOD EDUCATION TEST

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

The two-hour assessment contains 120 selected-response items⁴ covering five content areas: *Language and Literacy* (approximately 36 items), *Mathematics* (approximately 30 items), *Social Studies* (approximately 17 items), *Science* (approximately 17 items), and *Health and Physical Education, Creative and Performing Arts* (approximately 20 items).⁵ The reporting scale for the Praxis Early Childhood Education test ranges from 100 to 200 scaled-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.

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.

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.

⁴ Twenty of the 120 selected-response 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.

Panel 1 created a description of the just qualified candidate — the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, the panel first split into smaller groups to consider the just qualified candidate. The full panel then 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).

For Panel 2, the panelists began with the description of the just qualified candidate developed by Panel 1. Given that the multistate standard-setting study was designed to provide two recommendations for the same performance standard, it was important that panels use a consistent just qualified candidate description to frame their judgments. The panelists reviewed the just qualified candidate description, and any ambiguities were discussed and clarified.

PANELISTS' JUDGMENTS

The standard-setting process for the Praxis Early Childhood Education 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.

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 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 38 educators representing 16 states and Washington, D.C. (See Appendix A for a listing of panelists.) Nineteen panelists were teachers, eleven were college faculty, four were administrators or department heads, and four held other positions. All of the faculty members' job responsibilities included the training of early childhood 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)

	<i>N</i>	<i>%</i>
Current position		
Teacher	19	50
Administrator/Department head	4	11
College faculty	11	29
Other	4	11
Race		
White	27	71
Black or African American	6	16
Hispanic or Latino	1	3
Asian or Asian American	2	5
American Indian or Alaskan Native	1	3
Biracial	1	3
Gender		
Female	35	92
Male	3	8
Are you currently certified to teach this subject in your state?		
Yes	28	74
No	10	26
Are you currently teaching this subject in your state?		
Yes	30	79
No	8	21
Are you currently supervising or mentoring other teachers of this subject?		
Yes	27	71
No	11	29
At what K–12 grade level are you currently teaching this subject?		
Early Childhood (Birth – K)	6	16
Elementary (K–5 or K–6)	13	37
Other	5	13
Not currently teaching at the K–12 level	14	37

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

	<i>N</i>	<i>%</i>
Including this year, how many years of experience do you have teaching this subject?		
3 years or less	4	11
4–7 years	11	29
8–11 years	9	24
12–15 years	6	16
16 years or more	8	21
Which best describes the location of your K–12 school?		
Urban	8	21
Suburban	9	24
Rural	8	21
Not currently working at the K–12 level	13	34
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?		
Yes	11	29
Not college faculty	27	71

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	61.73	64.34
Lowest	44.85	54.85
Highest	78.95	81.45
SD	7.88	6.43
SEJ	1.76	1.51

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 Early Childhood Education test are 61.73 for Panel 1 and 64.34 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 — 62 for Panel 1 and 65 for Panel 2. The scaled scores associated with 62 and 65 raw points are 153 and 157, 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 Early Childhood Education test is 63.04 (out of a possible 100 raw-score points). The value was rounded to 64 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 64 raw points is 156.

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 CSEM above and below the recommended passing score are provided. The conditional standard error of measurement provided is an estimate and takes into account the recommended passing score and the number of questions on the test (see Lord, 1984).

Table 4***Passing Scores Within 1 and 2 CSEM of the Recommended Passing Score⁷***

Recommended passing score (CSEM)	Scale score equivalent
64 (4.82)	156
-2 CSEM	143
-1 CSEM	150
+ 1 CSEM	163
+ 2 CSEM	170

Note. CSEM = conditional standard error 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 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 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. Thirty-seven of the 38 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; 29 of the 38 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 38 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

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 scaled scores.

Thirty-five of the 38 panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 18 of the 38 panelists were *very comfortable*. Twenty-six of the 38 panelists indicated the recommended passing score was *about right*; 11 panelists indicated the passing score was *too low* with the remaining panelist indicating 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 Early Childhood Education 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 Early Childhood Education test, the recommended passing score⁸ is 64 out of a possible 100 raw-score points. The scaled score associated with a raw score of 64 is 156 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: Early Childhood Education (5025)*. 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.
- Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*, 239-243.
- 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

<u>Panelist</u>	<u>Affiliation</u>
Sheila Anderson	Weber State University (UT)
Nicholas Bellack	Professional Teaching Standards Board (WY)
Melissa Burnham	University of Nevada, Reno (NV)
Tina Burton	Ben Davis Elementary School (AL)
Monica Bybee	Heartland Elementary School (KY)
Brittany Courchesne	URI Child Development Center (RI)
Julie Davis	Eagle Grove Schools and William Penn University (IA)
Jennifer Ellis	Lincoln County Early Childhood (KY)
Karen Ferneding	Southern Arkansas University (AR)
Shauntra Franklin	Berewick Elementary School (NC)
Nicole A. Golden	Job Barnard Elementary School (DC)
Tiffany Harris-Greene	Mt. Laurel Board of Education (NJ)
Dana Heaton	Northport Elementary School (AL)
Shannon Hogue	Russell County School District (KY)
Angela Hunter	Buena Vista University (IA)
Genitia Johnson	Huntingtowne Farms Elementary School (NC)
Tammy Johnson	Richmond County Schools (NC)
C Morrell Jones	University of Arkansas, Monticello (AR)
Wendy Lundquist	Nashua School District (NH)
Christine Lux	Montana State University-Bozeman (MT)
Catherine Lyons	University of Nevada Las Vegas (NV)
Lindsey Maszk	Robert Place Miller (ND)

Participating Panelists With Affiliation (continued)

<u>Panelist</u>	<u>Affiliation</u>
Susan McGuire	River Vale Board of Education (NJ)
Anne McLaughlin	Community College of Baltimore County (MD)
Chloe Merrill	Weber State University (UT)
Andrea Newman	Anderson County Schools Sparrow Early Childhood Center (KY)
Elizabeth Park	Chaminade University of Honolulu (HI)
Allison Rackoff	Dale Avenue School (NJ)
Delia Richards	University of DC (DC)
Carrie Semmelroth	Boise State University (ID)
Tameka Shamery Rice	Taylorville Primary/Englewood Elementary School (AL)
Cheri Smith	Harding University (AR)
Jessica Sugrue	NH Department of Health and Human Services-Division for Children, Youth and Families-Child Development Bureau (NH)
Maureen Sweeney	Vernon Township Public Schools (NJ)
Jennifer Thomas	Swain County School System (NC)
Nanci Waterhouse	Salish Kootenai College (MT)
Lindsey Wright	Harford County Public Schools (MD)

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

APPENDIX B
STUDY AGENDA

AGENDA

Praxis Early Childhood Education (5025) Standard-Setting Study

Day 1

Welcome and Introduction

Overview of Standard Setting and the Praxis Early Childhood Education Test

Review the Praxis Early Childhood Education Test

Discuss the Praxis Early Childhood Education Test

Lunch

Define the Knowledge/Skills of a Just Qualified Candidate

Break

Standard-Setting Training

Round 1 Standard Setting Judgments

Collect Materials; End of Day 1

AGENDA

Praxis Early Childhood Education (5025) Standard-Setting Study

Day 2

Overview of Day 2

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 of the Just Qualified Candidate⁹

A just qualified candidate ...

Language and Literacy

1. Has foundational knowledge and recognition of the continuum of development for language and literacy in these areas: oral and written language, phonological awareness, concepts of print, fluency and comprehension
2. Recognizes that the role of literacy integrates across content areas and connects to children's real life experiences (e.g., literature, background)

Mathematics

1. Knows the developmental continuum of the foundational skills of math that include: number sense, operations, algebraic thinking, geometry, and measurement
2. Recognizes that there are different ways of showing mathematical knowledge through representation and interpretation

Social Studies

1. Knows the concepts of social and emotional development and its importance in shaping child development and learning
2. Knowledge of essential concepts of personal identity, cultural competence (responsiveness), civic participation in the environment and being able to integrate these across the curriculum and within the classroom environment

Science

1. Is familiar with scientific content and practices (e.g., inquiry methods), across all disciplines: physical science, earth and space science, life science, and engineering/technology
2. Can connect scientific knowledge to real life experiences

Health and Physical Education, Creative and Performing Arts

1. Knows the impact of health on learning and development across content areas
2. Knows the ways in which artistic practice informs, enriches, and complements teaching and learning
3. Knows the impact of physical development (esp. fine and gross motor development) on learning across content areas

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	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Current position				
Teacher	10	50	9	50
Administrator/Department head	2	10	2	11
College faculty	6	30	5	28
Other	2	10	2	11
Race				
White	15	75	12	67
Black or African American	2	10	4	22
Hispanic or Latino	1	5	0	0
Asian or Asian American	1	5	1	6
American Indian or Alaskan Native	1	5	0	0
Biracial	0	0	1	6
Gender				
Female	18	90	17	94
Male	2	10	1	6
Are you currently certified to teach this subject in your state?				
Yes	15	75	13	72
No	5	25	5	28
Are you currently teaching this subject in your state?				
Yes	17	85	13	72
No	3	15	5	28
Are you currently supervising or mentoring other teachers of this subject?				
Yes	16	80	11	61
No	4	20	7	39
At what K–12 grade level are you currently teaching this subject?				
Early Childhood (Birth – K)	4	20	2	11
Elementary (K–5 or K–6)	6	30	7	35
Other	3	15	2	11
Not currently teaching at the K–12 level	7	35	7	35

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

	Panel 1		Panel 2	
	<i>N</i>	%	<i>N</i>	%
Including this year, how many years of experience do you have teaching this subject?				
3 years or less	0	0	4	22
4–7 years	6	30	5	28
8–11 years	6	30	3	17
12–15 years	3	15	3	17
16 years or more	5	25	3	17
Which best describes the location of your K–12 school?				
Urban	4	20	4	22
Suburban	5	25	4	22
Rural	5	25	3	17
Not currently working at the K–12 level	6	30	7	39
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?				
Yes	6	30	5	28
Not college faculty	14	70	13	72

Table D2
Passing Score Summary by Round of Judgments

Panelist	Panel 1		Panel 2	
	Round 1	Round 2	Round 1	Round 2
1	57.85	59.45	55.75	54.85
2	62.80	59.70	70.90	70.40
3	66.05	65.60	67.70	65.00
4	73.25	73.10	76.35	81.45
5	79.90	78.95	57.20	64.95
6	68.05	65.85	56.95	58.95
7	70.05	67.60	60.60	60.30
8	55.25	52.75	66.65	63.85
9	54.30	54.70	61.45	63.90
10	59.65	61.25	61.00	63.00
11	62.70	60.35	63.70	65.00
12	55.75	58.90	55.75	61.25
13	57.35	61.20	57.65	55.75
14	62.25	63.15	61.40	61.35
15	55.70	53.70	71.60	69.55
16	52.30	54.50	53.10	57.90
17	72.70	72.90	70.60	69.65
18	45.30	44.85	71.00	71.00
19	65.80	64.90		
20	61.75	61.25		
Average	61.94	61.73	63.30	64.34
Lowest	45.30	44.85	53.10	54.85
Highest	79.90	78.95	76.35	81.45
SD	8.26	7.88	6.81	6.43
SEJ	1.85	1.76	1.60	1.51

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

	Strongly agree		Agree		Disagree		Strongly disagree	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• I understood the purpose of this study.	14	70	6	30	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	14	70	6	30	0	0	0	0
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	12	60	8	40	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	14	70	6	30	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	14	70	3	15	3	15	0	0
• The process of making the standard-setting judgments was easy to follow.	11	55	8	40	1	5	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	14	70	6	30	0	0		
• The between-round discussions	13	65	7	35	0	0		
• The knowledge/skills required to answer each test item	16	80	4	20	0	0		
• The passing scores of other panel members	1	5	15	75	4	20		
• My own professional experience	14	70	6	30	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?	6	30	11	55	3	15	0	0
	Too low		About right		Too high			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended passing score is:	8	40	11	55	1	5		

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

	Strongly agree		Agree		Disagree		Strongly disagree	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
• I understood the purpose of this study.	18	100	0	0	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	16	89	2	11	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	83	3	17	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	15	83	3	17	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	14	78	4	22	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	13	72	5	28	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	15	83	3	17	0	0		
• The between-round discussions	14	78	4	22	0	0		
• The knowledge/skills required to answer each test item	14	78	4	22	0	0		
• The passing scores of other panel members	5	28	11	61	2	11		
• My own professional experience	11	61	7	39	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?	12	67	6	33	0	0	0	0
	Too low		About right		Too high			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
• Overall, the recommended passing score is:	3	17	15	83	0	0		