



Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

**PRAXIS™ ENGLISH LANGUAGE ARTS: CONTENT KNOWLEDGE
(5038)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

March 2013

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ English Language Arts: Content Knowledge (5038) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING STATES

Panelists from 22 states, Washington, DC, and Guam were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either English teachers or college faculty who prepare English teachers and (b) familiarity with the knowledge and skills required of beginning English 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 English Language Arts: Content Knowledge test, the recommended passing score¹ is 79 out of a possible 110 raw-score points. The scaled score associated with a raw score of 79 is 167 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™ English Language Arts: Content Knowledge (5038) test, research staff from ETS designed and conducted a multistate standard-setting study² in March 2013 in Princeton, New Jersey. Education agencies³ recommended panelists with (a) experience as either English teachers or college faculty who prepare English teachers and (b) familiarity with the knowledge and skills required of beginning English teachers. Twenty-two states, Washington DC, and Guam (Table 1) were represented by 37 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating Jurisdictions and Number of Panelists

Alaska (2 panelists)	New Jersey (2 panelists)
Arkansas (1 panelist)	North Carolina (2 panelists)
Delaware (2 panelists)	North Dakota (2 panelists)
Guam (1 panelist)	Pennsylvania (1 panelist)
Hawaii (1 panelist)	Rhode Island (2 panelists)
Idaho (1 panelist)	South Carolina (1 panelist)
Kansas (1 panelist)	South Dakota (2 panelists)
Louisiana (1 panelist)	Tennessee (2 panelists)
Maine (1 panelist)	Utah (2 panelists)
Mississippi (2 panelists)	Washington, DC (1 panelist)
Montana (2 panelists)	Wisconsin (2 panelists)
Nevada (1 panelist)	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 jurisdiction, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in

² The multistate standard-setting study collected judgments for two related Praxis tests — Praxis English Language Arts: Content Knowledge (5038) and Praxis English Language Arts: Content and Analysis (5039). Separate technical reports were prepared for each test.

³ States and jurisdictions that currently use Praxis were invited to participate in the multistate standard-setting study.

accordance with applicable regulations. This study provides a recommended passing score,⁴ 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 Praxis English Language Arts: Content Knowledge 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 English Language Arts: Content Knowledge 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 English Language Arts: Content Knowledge 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.

⁴ In addition to the recommended passing score averaged across the two panels, the recommended passing scores for each panel are presented.

OVERVIEW OF THE PRAXIS ENGLISH LANGUAGE ARTS: CONTENT KNOWLEDGE TEST

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

The two and a half-hour assessment contains 130 selected-response items⁵ covering three content areas: *Reading* (approximately 49 items), *Language Use and Vocabulary* (approximately 33 items), and *Writing, Speaking and Listening* (approximately 48 items).⁶ The reporting scale for the Praxis English Language Arts: Content Knowledge 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.

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.

⁵ Twenty of the 130 selected-response items are pretest items and do not contribute to a candidate's score.

⁶ The number of items for each content area may vary slightly from form to form of the test.

DEFINING THE TARGET CANDIDATE

Following the review of the test, panelists described the target candidate. The *target 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 target 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 target candidate. The full panel then reconvened and, through whole-group discussion, created the description of the target candidate to use for the remainder of the study.

The written description of the target candidate summarized the panel discussion in a bulleted format. The description was not intended to describe all the knowledge and skills of the target 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 target candidate description).

For Panel 2, the panelists began with the description of the target 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 consistent target candidate description to frame their judgments. The panelists reviewed the target candidate description, and any ambiguities were discussed and clarified.

PANELISTS' JUDGMENTS

The standard-setting process for the Praxis English Language Arts: Content Knowledge 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 target 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 target candidate would answer the item correctly because the item is difficult for the target candidate. The higher the value, the more likely it is that the target 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 target candidate and the item and decided if, overall, the item would be difficult for the target candidate, easy for the target 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 target candidate are in the 0 to .30 range.
- Moderately difficult/easy items for the target candidate are in the .40 to .60 range.
- Easy items for the target 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 target 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 target 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 target 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 37 educators representing 22 states, Washington, DC, and Guam. (See Appendix A for a listing of panelists.) Nineteen panelists were teachers, thirteen were college faculty, four were administrators or department heads, and one held another position. Twelve of the thirteen faculty members’ job responsibilities included the training of English 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	51%
Administrator/Department head	4	11%
College faculty	13	35%
Other	1	3%
Race		
White	30	81%
Black or African American	4	11%
Asian or Asian American	1	3%
American Indian or Alaskan Native	1	3%
Native Hawaiian or Other Pacific Islander	1	3%
Gender		
Female	30	81%
Male	7	19%
Are you currently certified to teach this subject in your state?		
Yes	28	76%
No	9	24%
Are you currently teaching this subject in your state?		
Yes	29	78%
No	8	22%
Are you currently supervising or mentoring other teachers of this subject?		
Yes	22	59%
No	15	41%
At what K–12 grade level are you currently teaching this subject?		
Middle school (6–8 or 7–9)	2	5%
High school (9–12 or 10–12)	17	46%
Middle and High School	2	5%
Not currently teaching at the K–12 level	16	43%

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	5	14%
4–7 years	6	16%
8–11 years	9	24%
12–15 years	5	14%
16 years or more	12	32%
Which best describes the location of your K–12 school?		
Urban	7	19%
Suburban	6	16%
Rural	10	27%
Not currently working at the K–12 level	14	38%
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?		
Yes	12	32%
No	1	3%
Not college faculty	24	65%

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.⁷ 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).

⁷ 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	78.43	77.76
Lowest	69.00	69.95
Highest	90.25	88.20
SD	5.90	5.40
SEJ	1.35	1.27

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 English Language Arts: Content Knowledge test are 78.43 for Panel 1 and 77.76 for Panel 2 (out of a possible 110 raw-score points). The values were rounded to the next highest whole number, to determine the functional recommended passing score — 79 for Panel 1 and 78 for Panel 2. The scaled scores associated with 79 and 78 raw points are 167 and 166, 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 English Language Arts: Content Knowledge test is 78.10 (out of a possible 110 raw-score points). The value was rounded to 79 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 79 raw points is 167.

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 Score⁸***

Recommended passing score (CSEM)		Scale score equivalent
	79 (4.74)	167
-2 CSEMs	70	155
-1 CSEM	75	162
+ 1 CSEM	84	173
+ 2 CSEMs	89	180

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. All panelists *strongly agreed* or *agreed* that the standard-setting process was easy to follow.

All panelists reported that the description of the target candidate was at least *somewhat influential* in guiding their standard-setting judgments; 33 of the 37 panelists indicated the description was *very influential*. Thirty-six of the 37 panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. More than half of the panelists (twenty-four of the 37 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

⁸ 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.

All of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 22 of the 37 panelists were *very comfortable*. Thirty-three of the 37 panelists indicated the recommended passing score was *about right* with one of the remaining panelists indicating that the passing score was *too low* and three 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 English Language Arts: Content Knowledge 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 English Language Arts: Content Knowledge test, the recommended passing score⁹ is 79 out of a possible 110 raw-score points. The scaled score associated with a raw score of 79 is 167 on a 100–200 scale.

⁹ 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: English Language Arts: Content Knowledge (5038)*. 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

<u>Panelist</u>	<u>Affiliation</u>
Valentina Abordonado	Hawaii Pacific University (HI)
Jacqueline Bach	Louisiana State University (LA)
Amy Baker-Sheridan	Seaford Senior High School (DE)
Krista Bruggeman	Lennox High School (SD)
Jill Byrne	Beacon Charter High School for the Arts (RI)
Linda Constanzo Cahir	Kean University (NJ)
Granville Caldwell	Harding University High School (NC)
Sean Campbell	Homer High School (AK)
Danyka Davis	William G. Enloe High School (NC)
Anne Faulks	Appling Middle School (TN)
Creed Hansen	Sun Prairie High School (WI)
Heather Jo Harper	Century High School (ND)
A. Waller Hastings	West Liberty University (WV)
Patricia Hinchey	Penn State University (PA)
Thelma Hinds	Wilmington University (DE)
Peggy F. Hopper	Mississippi State University (MS)
Amanda Jackson	Nevada Virtual Academy (NV)
Kevin Jones	University of Arkansas Fort Smith (AR)
Laura S. Kim	Independence High School (TN)
Sherry Kinkopf	University of Southern Mississippi (MS)
Rachel Kittoe	West High School (AK)

Participating Panelists With Affiliation (continued)

<u>Panelist</u>	<u>Affiliation</u>
Debra Kohn	Smoky Valley High School (KS)
Gerri Lallo	Juanita Sanchez Education Complex (RI)
Wade Landsverk	Freedom High School (WI)
Liliana Maggioni	The Catholic University of America (DC)
Donna L. Miller	Aaniiih Nakoda College (MT)
Martina Nelson	Guam Department of Education JRMS (GU)
Stu Palmer	Mt. Ararat High School (ME)
Kathleen Rapp	Monmouth University (NJ)
Laura F. Scarpulla	Salt Lake City School District (UT)
Angela Schwer	Fairmont State University (WV)
Kari Lee Siko	Charleston Southern University (SC)
Juli Stricklan	Rigby High School (ID)
Abigail Tibbetts	Dawson County High School (MT)
Lorraine Wallace	Utah Valley University (UT)
Mary Weber	Hazen High School (ND)
Meghan Wounded Head	Hamlin High School (SD)

APPENDIX B
STUDY AGENDA

AGENDA

Praxis English Language Arts: Content Knowledge (5038) Standard-Setting Study

Day 1

Welcome and Introduction

Overview of Standard Setting and the Praxis English Language Arts: Content Knowledge Test

“Take” the Praxis English Language Arts: Content Knowledge Test (Take breaks as needed)

Discuss the Praxis English Language Arts: Content Knowledge Test

Define the Knowledge/Skills of a Target Candidate

Lunch

Define the Knowledge/Skills of a Target Candidate (continued)

Break

Standard-Setting Training for Selected-Response Items

Round 1 Standard Setting Judgments for Selected-Response

Collect Materials; End of Day 1

AGENDA

Praxis English Language Arts: Content Knowledge (5038) Standard-Setting Study

Day 2

Overview of Day 2

Standard Setting Training for Constructed-Response Questions¹⁰

Round 1 Standard Setting Judgments for Constructed-Response Questions

Round 1 Feedback & Round 2 Judgments

Break

Round 1 Feedback & Round 2 Judgments (continued)

Lunch

Round 1 Feedback & Round 2 Judgments (continued)

Break

Feedback on Round 2 Recommended Passing Score

Complete Final Evaluation

Collect Materials; End of Study

¹⁰ The multistate standard-setting study collected judgments for two related Praxis tests — Praxis English Language Arts: Content Knowledge (5038) and Praxis English Language Arts: Content and Analysis (5039). The Praxis English Language Arts: Content and Analysis (5039) test included two constructed-response (essay) items. Separate technical reports were prepared for each test.

APPENDIX C

TARGET CANDIDATE DESCRIPTION

Description of the Target Candidate¹¹

A target candidate ...

Reading Literature

1. Knows major works and authors of U.S., British, and world literature and can identify their historical, cultural and literary contexts
2. Understands the defining characteristics of primary literary genres and can identify the defining characteristics of major forms within each primary literary genre
3. Understands how textual evidence supports interpretations of a literary text
4. Understands how themes, literary elements and language contribute to the meaning of a text
5. Knows commonly used research based strategies for reading instruction and understands how reading strategies support comprehension

Informational Texts & Rhetoric

6. Understands how a variety of organizational patterns and text structures can be used to develop a central idea in informational texts
7. Understands rhetorical strategies that authors use to convey purpose and perspective in informational texts

Language Use and Vocabulary

8. Knows strategies for supporting language acquisition and vocabulary development (e.g., using affixes, decoding, word ladders, context)
9. Understands the conventions of Standard English grammar, usage, syntax, and mechanics

Writing, Speaking and Listening

10. Understands characteristics of clear and coherent writing and components of effective oral communication
11. Understands how awareness of mode, task, purpose, and audience contributes to effective written and oral communication
12. Knows commonly used research-based approaches to teaching and assessing reading, writing, speaking, and listening
13. Knows how to instruct students in effective use of digital media as a means of conducting research, enhancing communication and evaluating the credibility of sources
14. Knows how to adapt classroom instruction to accommodate various perspectives, cultures and backgrounds that students bring to speaking and writing

¹¹ Description of the target 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>N</i>	%
Current position				
Teacher	13	68%	6	33%
Administrator/Department head	3	16%	1	6%
College faculty	3	16%	10	56%
Other	0	0%	1	6%
Race				
White	16	84%	14	78%
Black or African American	1	5%	3	17%
Asian or Asian American	1	5%	0	0%
American Indian or Alaskan Native	0	0%	1	6%
Native Hawaiian or Other Pacific Islander	1	5%	0	0%
Gender				
Female	14	74%	16	89%
Male	5	26%	2	11%
Are you currently certified to teach this subject in your state?				
Yes	17	89%	11	61%
No	2	11%	7	39%
Are you currently teaching this subject in your state?				
Yes	17	89%	12	67%
No	2	11%	6	33%
Are you currently supervising or mentoring other teachers of this subject?				
Yes	9	47%	13	72%
No	10	53%	5	28%
At what K–12 grade level are you currently teaching this subject?				
Middle school (6–8 or 7–9)	1	5%	1	6%
High school (9–12 or 10–12)	12	63%	5	28%
Middle and High School	2	11%	0	0%
Not currently teaching at the K–12 level	4	21%	12	67%

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	2	11%	3	17%
4–7 years	4	21%	2	11%
8–11 years	4	21%	5	28%
12–15 years	4	21%	1	6%
16 years or more	5	26%	7	39%
Which best describes the location of your K–12 school?				
Urban	5	26%	2	11%
Suburban	4	21%	2	11%
Rural	7	37%	3	17%
Not currently working at the K–12 level	3	16%	11	61%
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?				
Yes	3	16%	9	50%
No	0	0%	1	6%
Not college faculty	16	84%	8	44%

Table D2
Passing Score Summary by Round of Judgments

Panelist	Panel 1		Panel 2	
	Round 1	Round 2	Round 1	Round 2
1	69.60	69.70	81.70	81.50
2	81.80	81.55	72.40	75.20
3	70.20	71.80	80.35	78.95
4	92.65	90.25	63.60	69.95
5	77.70	77.90	71.20	75.80
6	73.65	73.35	60.75	70.00
7	77.50	79.40	82.80	81.60
8	62.45	69.00	71.80	75.30
9	75.85	74.60	71.90	79.50
10	79.55	79.00	70.35	73.80
11	85.80	82.95	68.70	75.60
12	93.25	84.65	72.20	75.00
13	86.60	87.00	82.25	82.50
14	78.90	80.30	74.65	76.35
15	80.90	80.45	85.30	86.55
16	75.40	75.30	89.35	88.20
17	73.10	71.05	82.50	83.50
18	80.65	78.75	67.55	70.45
19	83.15	83.15		
Average	78.88	78.43	74.96	77.76
Lowest	62.45	69.00	60.75	69.95
Highest	93.25	90.25	89.35	88.20
SD	7.65	5.90	7.87	5.40
SEJ	1.76	1.35	1.86	1.27

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.	19	100%	0	0%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	19	100%	0	0%	0	0%	0	0%
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	16	84%	3	16%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	18	95%	1	5%	0	0%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	17	89%	2	11%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	17	89%	2	11%	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 target candidate	17	89%	2	11%	0	0%		
• The between-round discussions	9	47%	9	47%	1	5%		
• The knowledge/skills required to answer each test item	16	84%	2	11%	1	5%		
• The passing scores of other panel members	0	0%	17	89%	2	11%		
• My own professional experience	14	74%	5	26%	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	68%	6	32%	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:	0	0%	17	89%	2	11%		

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.	17	94%	1	6%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	13	72%	5	28%	0	0%	0	0%
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	16	89%	2	11%	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.	17	94%	1	6%	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 target candidate	16	89%	2	11%	0	0%		
• The between-round discussions	11	61%	7	39%	0	0%		
• The knowledge/skills required to answer each test item	16	89%	2	11%	0	0%		
• The passing scores of other panel members	4	22%	10	56%	4	22%		
• My own professional experience	10	56%	8	44%	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?	9	50%	9	50%	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:	1	6%	16	89%	1	6%		