Multistate Standard-Setting Technical Report

PRAXIS® MATHEMATICS (5165)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2021

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Mathematics (5165) test, research staff from Educational Testing Service (ETS) designed and conducted a distance-based multistate standard-setting study.

PARTICIPATING STATES

Panelists from 13 states and Washington, D.C., were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either secondary mathematics teachers or college faculty who prepare secondary mathematics teachers and (b) familiarity with the knowledge and skills required of beginning secondary mathematics 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* Mathematics test, the recommended passing score[1] is 39 out of a possible 60 raw-score points. The scale score associated with a raw score of 39 is 159 on a 100–200 scale.

[1] Results from the two panels participating in the study were averaged to produce the recommended passing score.

INTRODUCTION

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Mathematics (5165) test, research staff from ETS designed and conducted a distance-based multistate standard-setting study in January 2021. Education agencies[2] recommended panelists with (a) experience as either secondary mathematics teachers or college faculty who prepare secondary mathematics teachers and (b) familiarity with the knowledge and skills required of beginning secondary mathematics teachers. Thirteen states and Washington, D.C. (Table 1) were represented by 25 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating States, Washington, D.C., and Number of Panelists

Alabama (2 panelists)	Mississippi (2 panelists)
Arkansas (2 panelists)	Nevada (1 panelist)
Delaware (2 panelists)	South Carolina (2 panelists)
Idaho (1 panelist)	South Dakota (2 panelists)
Indiana (1 panelist)	Tennessee (2 panelists)
Kansas (2 panelists)	Washington, D.C. (2 panelists)
Maryland (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 and D.C., may want to consider the recommended passing score but also other sources of information when setting the final *Praxis* Mathematics passing score (see Geisinger & McCormick, 2010). A states and D.C., may accept the recommended passing score, adjust the score upward to reflect more stringent expectations, or

[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 passing scores for each panel are presented.

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 and 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* Mathematics test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a state and D.C., to recognize that any test score on any standardized test—including a *Praxis* Mathematics 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 state 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 does possess the required knowledge/skills. The state and D.C. need to consider which decision error is more important to minimize.

OVERVIEW OF THE *PRAXIS*® MATHEMATICS TEST

The Praxis® Mathematics Test at a Glance document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures the knowledge and competencies important for safe and effective beginning practice as a secondary school mathematics teacher.

The three-hour assessment contains 66 selected-response items[4] covering four content areas: *Number & Quantity and Algebra* (approximately 20 items), *Functions and Calculus* (approximately 20 items), *Geometry* (approximately 13 items), and *Statistics & Probability* (approximately 13 items).[5] The reporting scale for the *Praxis* Mathematics test ranges from 100 to 200 scale-score points.

PROCESSES AND METHODS

The design of the standard-setting study included two, independent expert panels of educators with experience with the test content and with new teachers or teacher candidates. Before the study, panelists received an email explaining the purpose of the standard-setting study and requesting that they review materials for the study, such as the test specifications and an overview presentation. This review helped familiarize the panelists with the general structure and content of the test. Additionally, panelists were asked to attend a brief, technology check meeting, to ensure that everyone could access the technology needed for the study.

For each panel, the first day of the standard-setting study began with a welcome by the meeting facilitator. After introductions of the panelists and ETS staff, the facilitator engaged the panel in a question and answer period about the overview presentation. Appendix B shows the agenda for the panel meeting.

REVIEWING THE TEST

Test familiarization was the first activity for the panel. The purpose of test familiarization is for the panelists to review the test and become familiar with the manner in which a candidate would take the test. After the facilitator described the purpose of the review and how to access the test[6], the standard-

[4] Six of the 66 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.
[6] The computer-administered test items were available through the ETS IBIS Content Review Tool.

setting panelists took the test and had a discussion of the content measured. This discussion helped bring the panelists to a shared understanding of what the test measures.

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. Overall, this discussion serves to reduce potential judgment errors later in the standard-setting process.

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 the final description of the just qualified candidate — the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. Each panel first worked separately by working in smaller and then a large group. Then both panels convened and, through whole-group discussion, combined the two descriptions in to the final version 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* Mathematics 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 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.

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 26 educators representing 12 states and D.C. (See Appendix A for a listing of panelists.) Twelve panelists were teachers, eight were college faculty, two were specialists, and three held another position. All of the faculty members' job responsibilities included the training of secondary mathematics 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	12	48
College faculty	8	32
Mathematics Specialist	2	8
Other	3	12
Race		
White	23	92
Black or African American	2	8
Gender		
Female	18	72
Male	7	28
Are you currently certified to teach mathematics in your state?		
Yes	22	88
No	3	12

(continues on next page)

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

	N	%
Are you currently teaching mathematics in your state?		
Yes	19	76
No	6	24
Are you currently supervising or mentoring mathematics teachers?		
Yes	23	92
No	2	8
At what K–12 grade level are you currently teaching mathematics?		
Elementary (K–5 or K–6)	1	4
Middle school (6–8 or 7–9)	1	4
Middle and High school	1	4
High school (9–12 or 10–12)	10	40
All Grades	1	4
Not currently teaching at the K–12 level	11	44
Including this year, how many years of experience do you have teaching mathematics?		
3 years or less	0	0
4–7 years	0	0
8–11 years	7	28
12–15 years	3	12
16 years or more	15	60
Which best describes the location of your K–12 school?		
Urban	3	12
Suburban	7	28
Rural	4	16
Not currently working at the K–12 level	11	44
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in mathematics?		
Yes	8	32
No	0	0
Not college faculty	17	68

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).

Table 3
Summary of Round 2 Standard-setting Judgments

	Panel 1	**Panel 2**
Average	37.18	39.94
Lowest	30.40	33.95
Highest	43.70	46.10
SD	4.22	3.64
SEJ	1.17	1.05

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* Mathematics test are 37.18 for Panel 1 and 39.94 for Panel 2 (out of a possible 60 raw-score points). The values were rounded to the next highest whole number, to determine the functional recommended passing score — 38 for Panel 1 and 40 for Panel 2. The scale scores associated with 38 and 40 raw points are 157 and 161, 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* Mathematics test is 38.56 (out of a possible 60 raw-score points). The value was rounded to 39 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 39 raw points is 159.

[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 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[8]

Recommended passing score (CSEM)		Scale score equivalent
39 (3.73)		159
-2 CSEM	32	143
-1 CSEM	36	152
+ 1 CSEM	43	168
+ 2 CSEM	47	177

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

FINAL EVALUATIONS

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

Panelists were also shown 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 just qualified candidate was at least *somewhat influential* in guiding their standard-setting judgments; 22 of the 25 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 (18 of the 25 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 scale scores.

All of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 22 of the 25 panelists were *very comfortable*. Twenty-four of the 25 panelists indicated the recommended passing score was *about right* 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* Mathematics 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* Mathematics test, the recommended passing score[9] is 39 out of a possible 60 raw-score points. The scale score associated with a raw score of 39 is 159 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®: The Praxis Study Companion: Mathematics (5165)*. 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**
Jodi Albers	Red Clay Consolidated School District (DE)
Holly Anthony	Tennesse Tech University (TN)
David Barnes II	Kansas State Department of Education (KS)
Stephen Bismarck	University of South Carolina Upstate (SC)
Sheila Blackmore	Bethany College (WV)
Tyesha Deas	A.C. Flora High School (SC)
Lacey Eckert	Sussex Technical High School (DE)
Ella Harris	Olathe Northwest High Sschool (KS)
Amanda Huffman	Pike High School (IN)
Paul Johanson	Brigham Young University - Idaho (ID)
Samantha Junkin	Auburn University at Montgomery (AL)
Melike Kara	Towson University (MD)
Ashley Kearney	Office of State Superintendent (DC)
Cindy Kroon	Montrose High School (SD)
Mary Martin	Middle Tennessee State University (TN)
Stephanie Marvel	Anne Arundel County Public Schools (MD)
Erin McCain	NW Arkansas Education Service Cooperative (AR)
Amanda Pendergrass	University of West Alabama (AL)
Adam Riazi	Cabell Midland High School (WV)
Amy Schander	Gayville-Volin High School (SD)
Thomas Schutt	DC Public Schools (DC)

(table continues)

Participating Panelists With Affiliation (continued)

Panelist	**Affiliation**
Sherra Shearer	Brandon High School (MS)
Douglas Speck	Southern Nevada Regional Professional Development Program (NV)
Rusty Young	Arkansas State University (AR)
Lauren Zarandona	Mississippi School for Math and Science (MS)

APPENDIX B

STUDY AGENDA

AGENDA

Praxis® Mathematics (5165)
Standard-Setting Study

Day 1

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Mathematics Test

Review the *Praxis* Mathematics Test

Discuss the *Praxis* Mathematics Test

Lunch

Define the Knowledge/Skills of a Just Qualified Candidate

Break

Define the Just Qualified Candidate *(continued)*

End of Day 1

AGENDA

Praxis® Mathematics (5165)
Standard-Setting Study

Day 2

Overview of Day 2

Define the Just Qualified Candidate *(continued)*

Standard-setting training presentation

Practice Round: Selected-response standard-setting judgments

Break

Practice Round: Data Discussion

Lunch Break

Round 1: Selected-response standard-setting judgments

Break

Round 1: Selected-response standard-setting judgments *(continued)*

End of Day 2

AGENDA

Praxis® Mathematics (5165)
Standard-Setting Study

Day 3

Overview of Day 3

Round 1 Feedback and Round 2 Judgments

Break

Round 1 Feedback and Round 2 Judgments *(continued)*

Break

Feedback on Round 2 Recommended Cut Score

Complete Final Evaluation

End of Study

APPENDIX C

JUST QUALIFIED CANDIDATE DESCRIPTION

Description of the Just Qualified Candidate[10]

A just qualified candidate…

Tasks of teaching mathematics across mathematical content areas

- Knows how to identify and reason about common mathematical misconceptions in student work
- Is familiar with identifying instructional items and examples that address a mathematical learning objective

Numbers & Quantity

1. Knows the structure and the basic operations and properties of the real and complex number systems.
2. Understands and is fluent with operations involving rational numbers
3. Understands how to determine the reasonableness of solutions within the context of a given problem
4. Understands ratios and proportions, especially in the context of dimensional analysis and estimation.
5. Knows properties of rational exponents and radicals as applied to number sets.

Algebra

6. Understands how to solve equations and inequalities using a variety of techniques such as graphical, algebraic, and tabular and understands how to justify the reasoning processes used.
7. Knows how varied techniques (e.g. graphical, algebraic, tabular) are used to solve systems of equations and inequalities
8. Knows how to find real and imaginary roots of common polynomials
9. Understands how to find and interpret the real and imaginary roots of quadratics
10. Understands how to rewrite algebraic expressions for specific purposes (e.g. factored form to find zeros, vertex form to find maxima or minima, point slope to slope intercept)
11. Knows how to model real world scenarios with algebraic expressions, including average rate of change

Functions

12. Understands how new functions are obtained from existing functions (e.g., compositions, transformations, and inverses)
13. Understands and can identify key characteristics of functions (e.g., domain, range, end behavior, increasing/decreasing/constant)
14. Understands how function behavior is analyzed using non-algebraic representations (e.g., graphs, mapping, and tables)
15. Understands how to solve basic trigonometric, logarithmic, and exponential equations
    Knows how to use basic trigonometric, logarithmic, and exponential expressions for modeling contextual situations.

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

Description of the Just Qualified Candidate (continued)

A just qualified candidate…

Calculus

16. Knows how to find the limit of a function numerically, algebraically or graphically.
17. Knows the derivative as a slope of a tangent line and as a rate of change
18. Is familiar with continuity and differentiability of functions.
19. Knows how and when to use standard differentiation and integration concepts

Geometry

20. Understands how trigonometry is applied to right triangles
21. Understands angle measurement in terms of radians and degrees.
22. Understands means for proving geometric properties (e.g., lines, angles, polygons, and their operations) using geometric and algebraic methods
23. Knows means for visualizing and reasoning algebraically among common 2D and 3D figures

Probability & Stats

24. Understands how to interpret a linear regression model (e.g., rate of change, intercepts, and correlation coefficient) in the context of the data
25. Understands and compute the concepts of interdependence and conditional probability (such as simple events, probabilities of compound events, conditional probabilities) and how to apply those concepts to data
26. Understands how to summarize, represent, and interpret common representations of qualitative and quantitative data
27. Knows how to use basic statistics to make inferences and informed decisions.
28. Is familiar with counting techniques such as permutations and combinations.

APPENDIX D

RESULTS

Table D1

Panel Member Demographics (by Panel)

	Panel 1		Panel 2	
	N	**%**	***N***	**%**
Current position				
Teacher	6	46	6	50
College faculty	4	31	4	33
Mathematics Specialist	1	8	1	8
Other	2	15	1	8
Race				
White	12	92	11	92
Black or African American	1	8	1	8
Gender				
Female	10	77	8	67
Male	3	23	4	33
Are you currently certified to teach mathematics in your state?				
Yes	12	92	10	83
No	1	8	2	17
Are you currently teaching mathematics in your state?				
Yes	11	85	8	67
No	2	15	4	33
Are you currently supervising or mentoring other mathematics teachers?				
Yes	12	92	11	92
No	1	8	1	8
At what K–12 grade level are you currently teaching this subject?				
Elementary (K–5 or K–6)	0	0	1	8
Middle school (6–8 or 7–9)	1	8	0	0
Middle and High school	0	0	1	8
High school (9–12 or 10–12)	6	46	4	33
All Grades	1	8	0	0
Not currently teaching at the K–12 level	5	38	6	50
Including this year, how many years of experience do you have teaching mathematics?				
3 years or less	0	0	0	0
4–7 years	0	0	0	0
8–11 years	3	23	4	33
12–15 years	1	8	2	17
16 years or more	9	69	6	50

(continued on next page)

Table D1 (continued)

Panel Member Demographics (by Panel)

	Panel 1		Panel 2	
	N	**%**	***N***	**%**
Which best describes the location of your K–12 school?				
Urban	2	15	1	8
Suburban	5	38	2	17
Rural	1	8	3	25
Not currently working at the K–12 level	5	38	6	50
If you are college faculty, are you currently involved in the training/preparation of teacher candidates in mathematics?				
Yes	4	31	4	33
No	0	0	0	0
Not college faculty	9	69	8	67

Table D2

Passing Score Summary by Round of Judgments

	Panel 1		Panel 2	
Panelist	**Round 1**	**Round 2**	**Round 1**	**Round 2**
1	33.90	33.80	42.45	42.20
2	38.70	37.25	35.80	37.55
3	35.50	35.35	39.15	40.90
4	41.40	39.30	44.75	43.50
5	43.40	42.75	44.80	42.30
6	40.50	38.50	40.35	40.35
7	33.55	34.25	33.05	33.95
8	30.00	33.80	38.15	38.70
9	28.60	30.40	36.45	38.05
10	30.10	32.30	32.90	34.10
11	40.70	40.30	40.65	41.55
12	42.75	41.65	51.45	46.10
13	45.20	43.70		
Average	37.25	37.18	40.00	39.94
Lowest	28.60	30.40	32.90	33.95
Highest	45.20	43.70	51.45	46.10
SD	5.62	4.22	5.36	3.64
SEJ	1.56	1.17	1.55	1.05

Table D3

Final Evaluation: Panel 1

	Strongly agree		Agree		Disagree		Strongly disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	12	92	1	8	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	12	92	1	8	0	0	0	0
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	11	85	2	15	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	11	85	2	15	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	12	92	1	8	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	10	77	3	23	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	92	1	8	0	0
• The between-round discussions	10	77	3	23	0	0
• The knowledge/skills required to answer each test item	11	85	2	15	0	0
• The passing scores of other panel members	7	54	4	31	2	15
• My own professional experience	10	77	3	23	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?	11	85	2	15	0	0	0	0

	Too low		About right		Too high	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0	12	92	1	8

Table D4

Final Evaluation: Panel 2

	Strongly agree		Agree		Disagree		Strongly disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	12	100	0	0	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	12	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.	12	100	0	0	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	12	100	0	0	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	10	83	2	17	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	12	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	10	83	2	17	0	0
• The between-round discussions	8	67	4	33	0	0
• The knowledge/skills required to answer each test item	11	92	1	8	0	0
• The passing scores of other panel members	5	42	6	50	1	8
• My own professional experience	8	67	4	33	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?	11	92	1	8	0	0	0	0

	Too low		About right		Too high	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0	12	100	0	0