Multistate Standard-Setting Technical Report for the *Praxis*® Elementary Education: Math Specialist (5037)

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

December 2021

Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Elementary Education: Math Specialist (5037) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

Participating States

Panelists from six states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either elementary education mathematics specialists or college faculty who prepare those elementary education mathematics specialists and (b) familiarity with the knowledge and skills required of beginning elementary education mathematics specialists.

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* Elementary Education: Math Specialist test, the recommended passing score is 36 out of a possible 60 raw-score points. The scale score associated with a raw score of 36 is 153 on a 100–200 scale.

Introduction

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Elementary Education: Math Specialist (5037) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in December 2021. Education agencies[1] recommended panelists with (a) experience as either elementary education mathematics specialists and (b) familiarity with the knowledge and skills required of beginning elementary education mathematics specialists. Six states (Table 1) were represented by 14 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating States and the Number of Panelists

Idaho (3 panelists)	Rhode Island (1 panelist)
Indiana (4 panelists)	Virginia (1 panelist)
Nevada (2 panelists)	West Virginia (3 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, 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, which represents the combined judgments of a group of experienced educators. Each state may want to consider the recommended passing score but also other sources of information when setting the final *Praxis* Elementary Education: Math Specialist passing score (see Geisinger & McCormick, 2010). A state 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 state's needs.

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

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* Elementary Education: Math Specialist test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows states to recognize that any test score on any standardized test—including a *Praxis* Elementary Education: Math Specialist test score—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows states to gauge the likelihood that the recommended passing score from the current 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 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 they should receive a license/certificate, but their 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 they should not receive a license/certificate, but they actually do possess the required knowledge/skills. States needs to consider which decision error is more important to minimize.

Overview of the *Praxis*® Elementary Education: Math Specialist Test

The *Praxis*® Elementary Education: Math Specialist *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level elementary education mathematics specialists have the knowledge/skills believed necessary for competent professional practice.

The two-hour assessment contains 75 selected-response items[2] covering two content areas: *Specialized Mathematics Knowledge for Teaching* (approximately 49 items) and *Pedagogical Knowledge and Instructional Leadership* (approximately 26 items).[3] The reporting scale for the *Praxis* Elementary Education: Math Specialist test ranges from 100 to 200 scale-score points.

Processes and Methods

The design of the standard-setting study included an expert panel. Before the study, panelists received an email explaining the purpose of the standard-setting study and requesting that they review the content specifications for the test. This review helped familiarize the panelists with the general structure and content of the test.

The standard-setting study began with a welcome and introduction by the meeting facilitator. The facilitator described the test, provided an overview of standard setting, and presented the agenda for the study. Appendix B shows the standard-setting study agenda.

Reviewing the Test

The standard-setting panelists first took the test and then discussed the content measured. This discussion helped bring the panelists to a shared understanding of what the test does and does not cover, which serves to reduce potential judgment errors later in the standard-setting process.

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 elementary education mathematics specialists or areas that address content particularly important for entry-level elementary education mathematics specialists.

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.

The panel created a description of the just-qualified candidate—the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panel

[2] Fifteen of the 75 selected-response items are pretest items and do not contribute to a candidate's score.
[3] The number of items for each content area may vary slightly from form to form of the test.

first split into smaller groups to consider the just-qualified candidate. Then they reconvened and, through whole-group discussion, determined 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 list 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* Elementary Education: Math Specialist test was a probability-based Modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). Using this method, 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 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 in the Modified Angoff method 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. Item-level data were highlighted to show when panelists converged in their judgments or diverged in their judgments. For the dichotomously-score items, this meant that at least two-thirds of the panelists' judgments were in the same difficulty range. For the constructed-response items, this meant that at least two-thirds of the panelists' judgments indicated the same score most likely earned by a just-qualified candidate.

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.

Results

Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panel included 14 educators representing six states. (See Appendix A for a listing of panelists.) Five panelists were teachers, three were college faculty, one was an administrator or department head, and five held other positions.

All of the educators were currently, or recently All of the faculty members' job responsibilities included the training of elementary education mathematics specialists.

Table 2
Panel Member Demographics

Background Survey Question	Number	Percent
What is your current position?	**N**	**%**
Teacher	5	36
Administrator or Department head	1	7
College faculty	3	21
Education Coordinator	5	36
How do you describe yourself (i.e., race/ethnicity)?	**N**	**%**
White	14	100
What is your gender?	**N**	**%**
Female	14	100
Are you currently certified to teach mathematics in your state?	**N**	**%**
Yes	13	93
No	1	7
Are you currently teaching mathematics in your state?	**N**	**%**
Yes	11	79
No	3	21
Are you currently supervising or mentoring other mathematics teachers?	**N**	**%**
Yes	11	79
No	3	21
At what P–12 grade level are you currently teaching this subject?	**N**	**%**
Middle school (6–8 or 7–9)	2	14
Elementary and Middle school	3	21
High school (9–12 or 10–12)	2	14
Not currently teaching at the P–12 level	7	50
Including this year, how many years of experience do you have teaching mathematics?	**N**	**%**
3 years or less	0	0
4–7 years	2	14
8–11 years	4	29
12–15 years	4	29
16 years or more	4	29

(table continues on the next page)

Table 2 (continued from the previous page)
Panel Member Demographics

Which best describes the location of your P–12 school?	**N**	**%**
Urban	1	7
Suburban	5	36
Rural	3	21
Not currently working at the P–12 level	5	36
If you are college faculty, are you currently involved in the training/ preparation of mathematics candidates?	**N**	**%**
Yes	3	21
No	0	0
Not college faculty	11	79

Standard-Setting Judgments

Table 3 summarizes the standard-setting judgments of each panelist and shows the passing score recommendations of each panelist at each round—the number of raw points needed to "pass" the test. The recommendations are the raw score points needed out of a maximum of 60.

Table 3
Raw Score Recommendation of Each Panelist by Round of Judgments

Panelist	Round 1	Round 2
1	38.00	36.80
2	34.60	36.40
3	35.10	34.90
4	35.45	34.20
5	34.60	33.60
6	38.25	37.15
7	33.90	35.30
8	41.55	38.05
9	33.70	34.70
10	33.35	33.25
11	34.10	34.90
12	32.90	33.30
13	39.00	39.00
14	31.70	34.20

Table 4 shows the summary statistics at each round of judgment. The mean represents the panel's passing score recommendation at each round. Table 4 also includes the standard deviations (SD) and the standard errors of judgment (SEJ) at each round. The SEJ is one way of estimating the reliability or consistency of a panel's standard-setting judgments. 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. (Appendix D provides the technical notes, which further describe the SEJ.)

Table 4

Summary Statistics by Round of Judgments

Statistic	Round 1	Round 2
Mean	35.44	35.41
Minimum	31.70	33.25
Maximum	41.55	39.00
SD	2.75	1.80
SEJ	0.73	0..48

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 (see Table 4).

The Round 2 mean score is the panel's final recommended passing score. The panel's passing score recommendation for the *Praxis* Elementary Education: Math Specialist test is 36 (out of a possible 60 raw-score points). The value was rounded to the next highest whole number, 36, to determine the functional recommended passing score. The scale score associated with 36 raw points is 153.

The conditional standard error of measurement (CSEM) around the recommended passing score is 3.83 raw points. A standard error represents the uncertainty associated with a test score (See Appendix D for further information about the CSEM.) Table 5 shows the raw scores and the scale scores associated with one and two CSEM below and above the recommended passing score.

Table 5

Scores 1 and 2 CSEM Around the Recommended Passing Score (RPS)

Scores	Raw Score Points out of 60	Praxis Scale Score Equivalent
RPS - 2 CSEM	29	137
RPS - 1 CSEM	33	146
RPS	**36**	**153**
RPS +1 CSEM	40	162
RPS +2 CSEM	44	171

Notes. CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 3.83 raw points. 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 then converted to scale scores.

Final Evaluations

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

Panelists were shown the panel's recommended passing score after Round 2 and asked, in the evaluation, (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 E.

All panelists *strongly agreed* or *agreed* that they understood the purpose of the study. All but one panelist *strongly agreed* or *agreed* that the facilitator's instructions and explanations were clear. All panelists *strongly agreed* or *agreed* that they were prepared to make their standard-setting judgments and 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. All but one of the panelists reported that round 2 discussions were at least *somewhat influential* in guiding their judgments. Ten of the 14 panelists indicated that their own professional experience was *very influential* in guiding their judgments.

Thirteen of the 14 panelists indicated they were *very comfortable* with the passing score they recommended; the remaining panelist was *somewhat comfortable* with the recommended passing score. All but one of the panelists indicated the recommended passing score was *about right*; one panelist felt 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* Elementary Education: Math Specialist 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* Elementary

Education: Math Specialist test, the recommended passing score is 36 out of a possible 60 raw-score points. The scale score associated with a raw score of 36 is 153 on a 100–200 scale.

References

Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education*, *17*, 59–88.

Brennan, R. L. (2002, October). Estimated standard error of a mean when there are only two observations (Center for Advanced Studies in Measurement and Assessment Technical Note Number 1). *Iowa City: University of Iowa*.

Cizek, G. J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

ETS. (in press). *The Praxis Series®: The Praxis Study Companion: Elementary Education: Math Specialist (5037)*. 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.

High, R. (2000). Dealing with 'outliers': How to maintain your data's integrity. *University of Oregon Computing News*, *15*(3), 14-16. Retrieved from http://hdl.handle.net/1794/3129

Lord. F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*, 239-283.

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.

Tannenbaum, R. J. (2011). Setting standards on *The Praxis Series*™ tests: A multistate approach. *R&D Connections*, *17*, 1-9.

Tannenbaum, R. J. (2012). A multistate approach to setting standards: An application to teacher licensure tests. *CLEAR Exam Review*, *23*(1), 18-24.

Appendix A: Panelists' Names & Affiliations

Participating Panelists With Affiliation and State

Panelist Name	**Panelists' Affiliation and State Abbreviation**
Veronica Blackham	Idaho State University (ID)
Sheila Blackmore	Bethany College (WV)
Hannah Bright	MSD of Martinsville (IN)
Kristina Caliri	WL Callahan School (RI)
Lori Eyth	Our Lady of Mt. Carmel (IN)
Angie Godfrey	Idaho State University (ID)
Nicole Jawhari	Southern Nevada Regional Professional Development Program (NV)
Stephanie Jones	Fairmont State University (WV)
Doris Mohr	University of Southern Indiana (IN)
Christin O'Keefe	Northwest Regional Professional Development Program/Washoe County School District (NV)
Marci Reddish	Regional Math Center (ID)
Candace Standley	AG Richardson Elementary School (VA)
Stephanie Stopiak	West Virginia University at Parkersburg (WV)
Amy Thomas	The Excel Center West (IN)

Appendix B: Agenda

Praxis® Elementary Education: Math Specialist (5037)

Standard-Setting Study

Day 1 Agenda

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Elementary Education: Math Specialist Test

Review the *Praxis* Elementary Education: Math Specialist Test

AM Break

Discuss the *Praxis* Elementary Education: Math Specialist Test

Lunch

Define the Knowledge/Skills of a Just-Qualified Candidate (small-group drafts)

PM Break

Define the Knowledge/Skills of a Just-Qualified Candidate (whole-group consensus)

Collect Materials; End of Day 1

Praxis® Elementary Education: Math Specialist (5037)

Standard-Setting Study

Day 2 Agenda

Overview of Day 2

Standard Setting Training in the Modified Angoff Method

AM Break

Round 1 Standard Setting Judgments

Lunch

Round 1 Feedback and Round 2 Judgments

PM Break

Feedback on Round 2 Recommended Passing Score

Complete Final Evaluation

Collect Materials; End of Study

Appendix C: Just-Qualified Candidate Description

Description of the Just-Qualified Candidate[4]

A just-qualified candidate…

I. Specialized Mathematics Knowledge for Teaching

1. Knows how to identify and connect multiple, commonly-used, representations to explain a mathematical concept.
2. Knows how to anticipate, identify, and appropriately address students' common misconceptions
3. Knows how to create opportunities for students to engage in utilizing the 8 Mathematical Practices (listed) within instructional activities to support a particular learning goal.
    1) Make sense of problems and persevere in solving them.
    2) Reason abstractly and quantitatively.
    3) Construct viable arguments and critique the reasoning of others.
    4) Model with mathematics.
    5) Use appropriate tools strategically.
    6) Attend to precision.
    7) Look for and make use of structure.
    8) Look for and express regularity in repeated reasoning.

II. Pedagogical Knowledge and Instructional Leadership

4. Knows how to use different instructional formats supports equitable learning of mathematics (e.g., whole-group, small-group, partner, individual).
5. Knows how to identify ways to provide each learner with opportunities to engage in productive struggle and to use multiple strategies (i.e., probing questions, critique other's reasoning. etc.) to assess and advance student thinking.
6. Is familiar with ways to support all learners (e.g., culturally diverse groups, ELLs) in the equitable learning of mathematics by embracing diversity in the classroom and school.
7. Knows multiple strategies for formatively and summatively assessing student learning, analyzing results to inform instruction, and providing targeted feedback to students and stakeholders.
8. Is familiar with how to effectively communicate and collaborate with different stakeholders (i.e., administrators, parents, community members, school-board members) regarding the use of data and best practices for math teaching and learning.

[4] 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…

9. Is familiar with coaching moves (e.g., telling, direct guidance, invitational guidance) and professional development resources to support teachers in reflecting on their own instructional practices and can provide research-based suggestions for improving instructional practices.

Mathematics Content

A. Numbers and Operations

10. Knows different developmental aspects of counting (e.g., cardinality, 1:1 correspondence, skip counting) and recognizes them as foundational for developing number sense.
11. Understands the basic concepts of fractions (part-whole, multiples of a unit fraction, as numbers, as division, as ratios and as percentages) and knows how to use different representations to model fractions and fraction operations.
12. Understands how to use basic math facts, number operations and properties (commutative, associative) to help students develop conceptual and procedural understanding.
13. Knows the impact of using precise vocabulary and conceptual math ideas for continuity in vertical alignment.
14. Understands place value in both whole numbers and decimals (conceptually and procedurally) and knows different representations to model place value.

B. Equations and Expressions

15. Understands how the equal sign and inequality signs are used to represent numeric relationships and can determine whether equations and inequalities are true.
16. Knows how to solve equations and inequalities using the four operations and can correctly apply the order of operations including real-world situations
17. Knows the use of properties of operations to evaluate and manipulate algebraic expressions and equations
18. Knows how to interpret mathematical models (e.g., graph, equation, table) in context.

C. Measurement and Geometry

19. Knows how to, conceptually and procedurally, apply geometry skills including multi-step applications of basic concepts (including area, perimeter, volume, angles/lines, shapes, etc.)
20. Knows how to use defining and non-defining attributes of two-dimensional and three-dimensional figures and to make and justify conjectures about geometric shapes and relations.
21. Knows how to select and use appropriate measurement tools for both standard and nonstandard units and can convert within a measurement system.

Description of the Just-Qualified Candidate (continued)

A just-qualified candidate…

D. Statistics

22. Knows how to interpret, analyze, and represent data in various forms and understands which form is most appropriate in a given situation
23. Knows how to solve problems involving measures of center (mean, median, mode) and measures of spread (range).

Appendix D: Technical Notes

Standard Error of Judgment (SEJ)

The standard error of judgment (SEJ) is one way of estimating the reliability or consistency of a panel's standard-setting judgments. 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 threshold score on the same form of the assessment. The 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 threshold scores (Tannenbaum & Katz, 2013).

The SEJ is calculated by dividing the standard deviation of the panelists' judgments (*SD*) by the square root of the number of panelists (*n*). The result serves as an estimate of the standard error of the mean (Brennan, 2002).

$$SEJ = \mathrm{SD} / \sqrt{n}$$

Conditional Standard Error of Measurement (CSEM)

The conditional standard error of measurement (*CSEM*) for a test is computed from the study value (*SV*) of the recommended passing score and the number of selected-response items (*n*) on the test (see Lord, 1984):

$$CSEM = \sqrt{(SV)(n - SV)/(n - 1)}$$

Appendix E: Final Evaluation Results

Table E1: Process Questions

Likert Statement	Strongly agree *N*	Strongly agree %	Agree *N*	Agree %	Disagree *N*	Disagree %	Strongly disagree *N*	Strongly disagree %
I understood the purpose of this study.	11	79	3	21	0	0	0	0
The instructions and explanations provided by the facilitators were clear.	9	64	4	29	1	7	0	0
The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	11	79	3	21	0	0	0	0
The explanation of how the recommended passing score is computed was clear.	11	79	3	21	0	0	0	0
The opportunity for feedback and discussion for round 2 judgments was helpful.	11	79	3	21	0	0	0	0
The process of making the standard-setting judgments was easy to follow.	9	64	5	36	0	0	0	0

Table E2: Final Evaluation: Standard-Setting Process

	Too much time *N*	Too much time %	About the right amount of time *N*	About the right amount of time %	Too little time *N*	Too little time %
Small group JQC drafts	0	0	13	93	1	7
Whole group JQC consensus	1	7	13	93	0	0
Training and practice for making standard-setting judgments	1	7	13	93	0	0
Round 1 judgments (independent)	3	21	11	79	0	0
Round 2 judgments (with discussion)	0	0	14	100	0	0

Table E3: Influences in Standard-Setting Judgments

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential *N*	Very influential %	Somewhat influential *N*	Somewhat influential %	Not influential *N*	Not influential %
The description of the just-qualified candidate	11	79	3	21	0	0
The round 2 discussion	2	14	11	79	1	7
The knowledge/skills required to answer each test item	8	57	6	43	0	0
The passing scores of other panel members	3	21	10	71	1	7
My own professional experience	10	71	4	29	0	0

Table E4: Comfort with the Panel's Recommendation

Question	Very comfort-able *N*	Very comfort-able %	Somewhat comfort-able *N*	Somewhat comfort-able %	Somewhat uncom-fortable *N*	Somewhat uncom-fortable %	Very uncom-fortable *N*	Very uncom-fortable %
Considering the process you followed, how comfortable are you with the panel's recommended cut score?	13	93	1	7	0	0	0	0

Table E5: Opinion of the Final Recommendation

Statement	Too low *N*	Too low %	About right *N*	About right %	Too high *N*	Too high %
Overall, the recommended passing score is:	0	0	13	93	1	7