Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

PRAXIS™ CORE ACADEMIC SKILLS FOR EDUCATORS
READING (5712)
WRITING (5722)
MATHEMATICS (5732)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2013

Copyright © 2013 by Educational Testing Service. All rights reserved. ETS, the ETS logo and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS) in the United States of America and other countries throughout the world. Praxis is a trademark of ETS.

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing passing scores (cut scores) for the Praxis™ Core Academic Skills for Educators Reading (5712), Writing (5722), and Mathematics (5732) subtests, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING JURISDICTIONS

Panelists from 23 states, Washington, DC, and Guam were recommended by their respective education agency. The education agencies recommended panelists with (a) experience preparing teachers candidates and (b) familiarity with the knowledge and skills required of candidates entering a teacher preparation program.

RECOMMENDED PASSING SCORES

ETS provides recommended passing scores from the multistate standard-setting study to help education agencies determine appropriate operational passing scores. For the Praxis Core Academic Skills for Educators subtests, the recommended passing scores[1] are

- **Reading**. The recommended passing score is 31 out of a possible 50 raw-score points. The scaled score associated with a raw score of 31 is 156 on a 100–200 scale.
- **Writing**. The recommended passing score is 44 out of a possible 70 raw-score points. The scaled score associated with a raw score of 44 is 162 on a 100–200 scale.
- **Mathematics**. The recommended passing score is 29 out of a possible 50 raw-score points. The scaled score associated with a raw score of 29 is 150 on a 100–200 scale.

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

To support the decision-making process of education agencies establishing passing scores (cut scores) for the Praxis™ Core Academic Skills for Educators Reading (5712), Writing (5722), and Mathematics (5732) subtests, research staff from ETS designed and conducted a multistate standard-setting study in January 2013 in Princeton, New Jersey. Education agencies[2] recommended panelists with (a) experience preparing teachers candidates and (b) familiarity with the knowledge and skills required of candidates entering a teacher preparation program. Twenty-three states, Washington, DC, and Guam (see Table 1) were represented by 74 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating Jurisdictions and Number of Panelists

Alaska (2 panelists)	North Dakota (2 panelists)
Arkansas (4 panelists)	Nebraska (4 panelists)
Connecticut (4 panelists)	New Hampshire (3 panelists)
Delaware (3 panelists)	New Jersey (3 panelists)
Guam (4 panelists)	Nevada (3 panelists)
Hawaii (3 panelists)	Rhode Island (2 panelists)
Iowa (1 panelist)	South Carolina (4 panelists)
Kentucky (2 panelists)	Tennessee (2 panelists)
Louisiana (4 panelists)	Vermont (2 panelists)
Maryland (4 panelists)	Washington, DC (1 panelist)
Maine (4 panelists)	Wisconsin (2 panelists)
Mississippi (4 panelists)	West Virginia (4 panelists)
North Carolina (3 panelists)	

The following technical report contains three sections. The first section describes the content and format of each subtest. The second section describes the standard-setting processes and methods. The third section presents the results of the standard-setting study.

ETS provides recommended passing scores 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 scores for

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

each subtest in accordance with applicable regulations. This study provides recommended passing scores, which represents the combined judgments of two panels of experienced educators. Each jurisdiction may want to consider the recommended passing scores and other sources of information when setting the final Praxis Core Academic Skills for Educator passing scores (see Geisinger & McCormick, 2010). A jurisdiction may accept the recommended passing scores, adjust one or more of the scores upward to reflect more stringent expectations, or adjust one or more of the scores downward to reflect more lenient expectations. There are no *correct* decisions; 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 scores are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the Praxis Core Academic Skills for Educators subtests and the latter, the reliability of panelists' passing-score recommendations. The SEM allows a jurisdiction to recognize that any test score on any standardized test—including the Praxis Core Academic Skills for Educator subtests—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 allow a jurisdiction to gauge the likelihood that a 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 jurisdiction should consider the likelihood of classification error. 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 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 may be more important to minimize.

OVERVIEW OF THE PRAXIS CORE ACADEMIC SKILLS FOR EDUCATORS

The Praxis Core Academic Skills for Educators *Test at a Glance* documents (ETS, in press) describes the purpose and structure of each subtest. In brief, the Praxis Core Academic Skills for Educators subtests measure whether candidates entering a teacher preparation program have the necessary reading, writing, and mathematical knowledge/skills. Each subtest — Reading, Writing, and Mathematics — is administered and scored separately.

- **Reading.** The 85-minute subtest contains 56 multiple-choice items[3] covering three content areas: *Key Ideas and Details*, *Craft, Structure and Language Skills*, and *Integration of Knowledge and Ideas*.
- **Writing.** The 100-minute subtest contains 40 multiple-choice items[4] covering two content areas: *Text Types, Purposes and Production* and *Language and Research Skills for Writing* .
- **Mathematics.** The 85-minute subtest contains 56 multiple-choice and numeric-entry items[5] covering four content areas: *Number and Quantity*, *Algebra and Functions*, *Geometry*, and *Statistics an Probability*.

The reporting scale for all three Praxis Core Academic Skills for Educators subtests ranges from 100 to 200 scaled-score points.

[3] Six of the 56 multiple-choice items are pretest items and do not contribute to a candidate's score.
[4] Six of the 40 multiple-choice items are pretest items and do not contribute to a candidate's score.
[5] Six of the 56 multiple-choice or numeric entry items are pretest items and do not contribute to a candidate's score.

PROCESSES AND METHODS

The design of the standard-setting study included multiple expert panels. The first panel (Panel 1) reviewed each of the three subtests. The remaining three panels (Panels 2, 3 and 4) each reviewed one of the subtests. Thus, each subtest was reviewed by 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 each subtest. This review helped familiarize the panelists with the general structure and content of the Reading, Writing, and Mathematics subtests.

For each panel, the standard-setting study began with a welcome and introduction by the meeting facilitator. The facilitator described the Praxis Core Academic Skills for Educators (all three subtests), provided an overview of standard setting, and presented the agenda for the study. Appendix B shows the agenda for Panel 1. (Agendas for Panels 2, 3 and 4 were variations depending on the subtest being reviewed.)

REVIEWING THE SUBTEST

The standard-setting panelists first took the particular subtest (Reading, Writing, or Mathematics) and then discussed it.[6] This discussion helped to bring the panelists to a shared understanding of what the subest 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 subtest. Panelists were asked to remark on any content areas that would be particularly challenging for candidates entering a teacher preparation program and areas that address content that would be particularly important for candidates entering a teacher preparation program.

[6] Panel 1 reviewed each of the three subtests; Panels 2, 3 and 4 reviewed one of the subtests for the purpose of the standard-setting study. Panels 2, 3 and 4 reviewed a second subtest as part of a research project (not included in the recommended passing scores).

Defining the Target Candidate

Following the review of the subtest, 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 subtest 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 descriptions for each of the three subtests).

For Panels 2, 3 and 4, 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 descriptions 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 Core Academic Skills for Educators was conducted separately for each subtest. Two subtests — Reading and Mathematics — include only dichotomously-scored items (multiple-choice items for Reading and multiple-choice and numeric-entry items for Mathematics). The Writing subtest includes both dichotomously-scored (multiple-choice items) and constructed-response (essays) items. Panelists received training in two distinct standard-setting approaches: one standard-setting approach for the dichotomously-scored items contained on all three subtests and another approach for the constructed-response items on the Writing subtest.

A panel's passing score for the Reading or Mathematics subtests is based on passing scores recommended by panelists for the dichotomously-scored items. For the Writing subtest, a panel's

passing score is the sum of the interim passing scores recommended by the panelists for (a) the dichotomously-scored items and (b) the essays. As with scoring and reporting, the panelists' judgments for the essays were weighted such that they contributed approximately 51% of the overall Writing score.

Dichotomously scored items. The standard-setting process for the multiple-choice items on the Reading, Writing, and Mathematics subtests, as well as the numeric-entry items on the Mathematics subtest, 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.

Constructed-response (essay) items. An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for the constructed-response (essay) items on the Writing subtest.

For this portion of the study, a panelist decided on the assigned score value that would most likely be earned by the target candidate for each essay. Panelists were asked first to review the definition of the target candidate and then to review the essay and its rubric. The rubric for an essay defines (holistically) the quality of the evidence that would merit a response earning a particular score. During this review, each panelist independently considered the level of knowledge/skill required to respond to the essay and the features of a response that would earn a particular score, as defined by the rubric. Each panelist decided on the score most likely to be earned by the target candidate from the possible values a test taker can earn.

A test-taker's response to an essay on the Writing subtest is independently scored by two raters, and the sum of the raters' scores is the assigned score[7]; possible scores, therefore, range from two (both raters assigned a score of one) to twelve (both raters assigned a score of six). For their ratings, each panelist decided on the score most likely to be earned by a target candidate from the following possible values: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12. For each of the essays, panelists recorded the score (2 through 12) that a target candidate would most likely earn.

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.

Multiple Rounds. 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. For dichotomously-scored items, 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.

[7] If the two raters' scores differ by more than one point (non-adjacent), the Chief Reader for that item assigns the score, which is then doubled.

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 Panels 2, 3 and 4. The item-level judgments and resulting discussions for Panels 2, 3 and 4 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 panels included 74 educators representing 23 states, Washington, DC, and Guam. (See Appendix A for a listing of panelists.) In brief, 53 were college faculty, 15 were administrators or department heads, and six held another position. All but one faculty members' job responsibilities included training of teacher candidates.

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

Table 2

Panel Member Demographics (Across Panels)

	N	%
Current position		
College faculty	53	72%
Administrator/Department head	15	20%
Other	6	8%
Race		
White	56	76%
Black or African American	10	14%
Hispanic or Latino	3	4%
Asian or Asian American	1	1%
Native Hawaiian or Other Pacific Islander	2	3%
Other	2	3%
Gender		
Female	50	68%
Male	24	32%
Are you currently supervising or mentoring beginning teachers?		
Yes	63	85%
No	11	15%
Are you currently involved in the training/preparation of teacher candidates?		
Yes	73	99%
No	1	1%
Not college faculty	0	0%
In all, how many years have you taught teacher candidates?		
3 years or less	11	15%
4 - 7 years	17	23%
8 - 11 years	15	20%
12 - 15 years	15	20%
16 years or more	16	22%

STANDARD-SETTING JUDGMENTS

Tables 3-5 summarize the standard-setting judgments (Round 2) of panelists. The tables also include 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.[8] 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 subtest. For each of the subtests, 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 (see Tables D2 and D3).

Table 3
Summary of Round 2 Standard-setting Judgments – Reading

	Panel 1	**Panel 2**
Average	30.62	30.23
Lowest	22.50	22.10
Highest	37.35	37.00
SD	3.49	4.42
SEJ	0.71	1.07

Table 4
Summary of Round 2 Standard-setting Judgments – Writing

	Panel 1	**Panel 3**
Average	42.99	43.73
Lowest	37.80	34.20
Highest	48.75	54.80
SD	2.90	4.48
SEJ	0.59	1.06

[8] 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 5
Summary of Round 2 Standard-setting Judgments – Mathematics

	Panel 1	Panel 4
Average	26.92	29.59
Lowest	19.05	25.60
Highest	32.20	34.30
SD	3.20	3.00
SEJ	0.65	0.80

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 subtest and for each panel (see Tables D2 and D3 in Appendix D). The Round 2 average score is the panel's recommended passing score.

Reading. The panels' passing score recommendations for the Praxis Core Academic Skills for Educator Reading subtest are 30.62 for Panel 1 and 30.23 for Panel 2 (out of a possible 50 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended passing scores — 31 for both Panel 1 and Panel 2. The scaled score associated with 31 raw points is 156.

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 for the Reading subtest. The panels' average passing score recommendation for the Reading subtest is 30.43 (out of a possible 50 raw-score points). The value was rounded to 31 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 31 raw points is 156.

Table 6 presents the estimated conditional standard error of measurement (CSEM) around the recommended passing score for the Reading subtest. A standard error represents the uncertainty associated with a subtest 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 6

Passing Scores Within 1 and 2 CSEMs of the Recommended Passing Score: Reading[9]

Recommended passing score (CSEM)		Scale score equivalent
31 (3.47)		156
- 2 CSEMs	25	140
-1 CSEM	28	148
+1 CSEM	35	166
+ 2 CSEMs	38	174

Note. **CSEM = conditional standard error of measurement.**

Writing. The panels' passing score recommendations for the Praxis Core Academic Skills for Educator Writing subtest are 42.99 for Panel 1 and 43.73 for Panel 3 (out of a possible 70 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended passing scores — 43 for Panel 1 and 44 for Panel 3. The scaled scores associated with 43 and 44 raw points are 160 and 162, 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 for the Writing subtest. The panels' average passing score recommendation for the Writing subtest is 43.36 (out of a possible 70 raw-score points). The value was rounded to 44 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 44 raw points is 162.

Table 7 presents the estimated conditional standard error of measurement (CSEM) around the recommended passing score for the Writing subtest. A standard error represents the uncertainty associated with a subtest 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.

[9] 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.

Table 7

Passing Scores Within 1 and 2 CSEMs of the Recommended Passing Score: Writing[10]

Recommended passing score (CSEM)		Scale score equivalent
44 (3.90)		162
- 2 CSEMs	37	150
-1 CSEM	41	158
+1 CSEM	48	170
+ 2 CSEMs	52	176

Note. **CSEM = conditional standard error of measurement.**

Mathematics. The panels' passing score recommendations for the Praxis Core Academic Skills for Educator Mathematics subtest are 26.92 for Panel 1 and 29.59 for Panel 4 (out of a possible 50 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended passing scores — 27 for Panel 1 and 30 for Panel 4. The scaled scores associated with 27 and 30 raw points are 146 and 154, 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 for the Mathematics subtest. The panels' average passing score recommendation for the Mathematics subtest is 28.26 (out of a possible 50 raw-score points). The value was rounded to 29 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 29 raw points is 150.

Table 8 presents the estimated conditional standard error of measurement (CSEM) around the recommended passing score for the Mathematics subtest. A standard error represents the uncertainty associated with a subtest 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.

[10] 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.

Table 8

Passing Scores Within 1 and 2 CSEMs of the Recommended Passing Score: Mathematics[11]

Recommended passing score (CSEM)		Scale score equivalent
29 (3.53)		150
- 2 CSEMs	22	132
-1 CSEM	26	142
+1 CSEM	33	162
+ 2 CSEMs	37	172

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 scores.

Panelists were also shown their panel's recommended passing score for the subtest(s) they reviewed 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 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; 49 of the 74 panelists indicated the description was *very influential*. All but one of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Approximately three-quarters of the panelists (52 of the 74 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

[11] 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.

- **Reading.** Across Panels 1 and 2, all but one of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 29 of the 41 panelists were *very comfortable*. Forty of the 41 panelists indicated the recommended passing score was *about right* with the remaining panelist indicating that the passing score was *too high*.
- **Writing.** Across Panels 1 and 3, 38 of the 42 panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 25 of the 42 panelists were *very comfortable*. Thirty-eight of the 42 panelists indicated the recommended passing score was *about right* with the remaining four panelists indicating that the passing score was *too high*.
- **Mathematics.** Across Panels 1 and 4, all but two of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 30 of the 39 panelists were *very comfortable*. Thirty-six of the 39 panelists indicated the recommended passing score was *about right* with the remaining three panelists indicating that the passing score was *too low*.

SUMMARY

To support the decision-making process of education agencies establishing passing scores (cut scores) for the Praxis Core Academic Skills for Educators Reading, Writing, and Mathematics subtests, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

ETS provides recommended passing scores from the multistate standard-setting study to help education agencies determine appropriate operational passing scores. For the Praxis Core Academic Skills for Educators subtests, the recommended passing scores[12] are

- **Reading**. The recommended passing score is 31 out of a possible 50 raw-score points. The scaled score associated with a raw score of 31 is 156 on a 100–200 scale.
- **Writing**. The recommended passing score is 44 out of a possible 70 raw-score points. The scaled score associated with a raw score of 44 is 162 on a 100–200 scale.
- **Mathematics**. The recommended passing score is 29 out of a possible 50 raw-score points. The scaled score associated with a raw score of 29 is 150 on a 100–200 scale.

[12] Results from two panels participating in the study were averaged to produce the recommended passing score for each subtest.

REFERENCES

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

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: Core Academic Skills for Educators – Reading (5712).* Princeton, NJ: Author.

ETS. (in press). *The Praxis Series: Core Academic Skills for Educators – Writing (5722)*. Princeton, NJ: Author.

ETS. (in press). *The Praxis Series: Core Academic Skills for Educators – Mathematics (5732)*. 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.

Hambleton, R. K., & Plake, B.S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41-55.

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**
Travis Allen	Husson University (ME)
Gwendolyn Autin	Southeastern Louisiana University (LA)
Karen Berard-Reed	Rhode Island College (RI)
Gina Bittner	Peru State College (NE)
Simone Bollinger	Guam Community College (GU)
Amy D. Broemmel	University of Tennessee – Knoxville (TN)
Gary Bunn	University of Central Arkansas (AR)
Debbie Bush	West Virginia Wesleyan (WV)
Roland Caron	University of Maine at Fort Kent (ME)
LaShundia Carson	Alcorn State University (MS)
Agnes Cave	The Catholic University of America (DC)
Anita S. Charles	Bates College (ME)
John Ciochine	University of New Haven (CT)
Cecil Clark	Delaware State University (DE)
Debra Coventry	Henderson State University (AR)
Allison Swan Dagen	West Virginia University (WV)
Jane Dalton	Maine College of Art (ME)
Mark Dewalt	Winthrop University (SC)
John Doak	University of Arkansas – Fort Smith (AR)
Sara Eisenhardt	Northern Kentucky University (KY)
Valjeaner Ford	University of North Carolina at Pembroke (NC)
Robin Kesterson Franklin	Charleston Southern University (SC)
Daphne Ghorbani	University of Mary (ND)
Cyndi Giorgis	University of Nevada Las Vegas (NV)
Adam Goldberg	Southern Connecticut State University (CT)
Donna Grace	University of Hawaii at Manoa (HI)
Anthony Graham	North Carolina Agricultural and Technical State University (NC)
Jo Hoffman	Kean University (NJ)

Participating Panelists With Affiliation

Panelist	**Affiliation**
Gilda M. Jones	Southern University at New Orleans (LA)
Jennifer Jordan	University of Tennessee (TN)
Ute Kaden	University of Alaska Fairbanks (AK)
Cindy Leonard	University of Vermont (VT)
Emily Lin	University of Nevada, Las Vegas (NV)
Frank Livoy	University of Delaware (DE)
Calandra D. Lockhart	University of Charleston (WV)
Jeffrey L. Lofthus	University of Alaska Southeast (AK)
Rexton Lynn	Ramapo College of New Jersey (NJ)
Elaine McClure	Morris College (SC)
Debra Poole Miller	Stevenson University (MD)
Mary Jane Miller	University of Guam (GU)
Margaret Mize	Chaminade University of Honolulu (HI)
Lori Navarrete	Nevada State College (NV)
Linda Neuzil	University of Pikeville (KY)
Kathleen Norris	Plymouth State University (NH)
Norman Norris	Nicholls State University (LA)
Michael A. Nugent	University of Maryland Eastern Shore (MD)
Alan Olson	Valley City State University (ND)
Janet Painter	Lenoir-Rhyne University (NC)
Margaret Queenan	University of Bridgeport (CT)
Elaine Razzano	Lyndon State College (VT)
Juliette Relihan	Salve Regina University (RI)
Kathryn (Kass) Rempp	Hastings College (NE)
Karen Rigoni	University of Wisconsin – Milwaukee (WI)
David Roloff	The University of Wisconsin – Stevens Point (WI)
Kim Rotruck	Frostburg State University (MD)

Participating Panelists With Affiliation

Panelist	**Affiliation**
Scott Rozell	iteachHawaii (HI)
Annette Rycharski	Felician College (NJ)
John Sanchez	University of Guam (GU)
Michelle M. S. Santos	University of Guam (GU)
Jesse Sealey	Chadron State College (NE)
Donna Shea	Mississippi State University (MS)
Patricia R. Sherblom	Wesley College (DE)
Ron Siers, Jr.	Salisbury University (MD)
Susan Stachowski	Mississippi Community College (MS)
Tanya Sturtz	Keene State College (NH)
Darrell Thompson	Bluefield State College (WV)
Loretta Walton-Jaggers	Grambling State University (LA)
Maureen Ward	SAU 18 (NH)
Karen C. Waters	Sacred Heart University (CT)
Elizabeth A. Wells	The University of Arkansas at Pine Bluff (AR)
Kelly Welsh	University of Nebraska – Omaha (NE)
Dennis Williams	Jackson State University (MS)
Reginald Harrison Williams	South Carolina State University (SC)
Barbara C. Wilt	Morningside College (IA)

APPENDIX B

STUDY AGENDA

AGENDA

Praxis™ Core Academic Skills for Educators (5712, 5722 & 5732) Standard-Setting Study

Day 1

Welcome and Introductions

Overview of Standard Setting and the Praxis "Core" Test

"Take" the Praxis "Core" Test: Reading
(Take breaks as needed)

Discuss the Praxis "Core" Test: Reading

Break

Who is the Just Qualified Candidate (JQC)?

Define the Knowledge/Skills of a JQC: Reading

Lunch

Standard-Setting Training for Multiple-Choice Items

Round 1 Standard-Setting Judgments: Reading

Break

Round 1 Feedback and Round 2 Judgments: Reading

Collect Materials; End of Day 1

AGENDA

Praxis™ Core Academic Skills for Educators (5712, 5722 & 5732) Standard-Setting Study

Day 2

Overview of Day 2

"Take" the Praxis "Core" Test: Writing
(Take breaks as needed)

Discuss the Praxis "Core" Test: Writing

Define the Knowledge/Skills of a JQC: Writing

Break

Review Standard-Setting for Multiple Choice Items & Standard Setting Training for Constructed-Response Items

Round 1 Standard-Setting Judgments: Writing

Lunch

Round 1 Feedback and Round 2 Judgments: Writing

"Take" the Praxis "Core" Test: Mathematics
(Take breaks as needed)

Discuss the Praxis "Core" Test: Mathematics

Collect Materials; End of Day 2

AGENDA

Praxis™ Core Academic Skills for Educators (5712, 5722 & 5732) Standard-Setting Study

Day 3

Overview of Day 3

Define the Knowledge/Skills of a JQC: Mathematics

Review Standard Setting for Multiple Choice Items

Round 1 Standard-Setting Judgments: Mathematics

Lunch

Round 1 Feedback and Round 2 Judgments: Mathematics

Break

Feedback on Round 2 Recommended Passing Score

Complete Final Evaluation

Collect Materials; End of Study

APPENDIX C

TARGET CANDIDATE DESCRIPTIONS

Description of the Target Candidate[13]

A target candidate …

Reading:

1. Can infer logically from an informational text
2. Can identify specific details of text including how and why individuals, events or ideas interact as well as drawing inferences or making implications
3. Can summarize central ideas, themes, and key details
4. Can identify text organization in terms of cause/effect, compare/contrast, problem/solution, and fact/opinion
5. Can identify author's point of view
6. Can apply knowledge and use of language (vocabulary multiple meanings, use of context clues) to comprehend when reading
7. Knows how evidence relates to an argument within a text
8. Can apply ideas from reading to draw conclusions and make predictions

Writing:

1. Organize and develop ideas logically, making coherent connections and supporting with appropriate details
2. Establish clear theses (i. e., focus)
3. Use effective sentence structures to strengthen writing
4. Effectively edit sentences for grammar, conventions, and usage
5. Effectively revise sentences for style and clarity
6. Identify information and credit sources that are relevant and credible to a particular research topic

Mathematics:

1. Use ratios and proportions to solve real world problems
2. Understand mathematical operations with rational numbers (e.g., whole numbers, negative numbers, decimals and fractions)
3. Solve real-world and mathematical problems using geometric concepts (e.g., angles, area, volume) and applying basic formulas
4. Use numerical and algebraic expressions to solve real-world and mathematical problems
5. Solve one-variable equations and inequalities
6. Analyze equations and graphical representations of real-world problems
7. Can summarize (central tendency, variability), represent and interpret data
8. Understand sampling and probability to evaluate outcomes
9. Understands the relationship between two sets of data

[13] 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	
	N	**%**	**N**	**%**
Current position				
College faculty	18	75%	10	59%
Administrator or Department head	5	21%	5	29%
Other	1	4%	2	12%
Race				
White	17	71%	13	76%
Black or African American	4	17%	3	18%
Hispanic or Latino	0	0%	0	0%
Asian or Asian American	1	4%	0	0%
Native Hawaiian or Other Pacific Islander	1	4%	1	6%
Other	1	4%	0	0%
Gender				
Female	19	79%	9	53%
Male	5	21%	8	47%
Are you currently supervising or mentoring beginning teachers?				
Yes	20	83%	14	82%
No	4	17%	3	18%
Are you currently involved in the training/preparation of teacher candidates?				
Yes	23	96%	17	100%
No	1	4%	0	0%
In all, how many years have you taught teacher candidates?				
3 years or less	4	17%	3	18%
4 - 7 years	3	13%	6	35%
8 - 11 years	5	21%	3	18%
12 - 15 years	9	38%	2	12%
16 years or more	3	13%	3	18%

Table D1 (continued)

Panel Member Demographics (By Panel)

	Panel 3		Panel 4	
	N	**%**	**N**	**%**
Current position				
College faculty	15	83%	10	67%
Administrator or Department head	1	6%	4	27%
Other	2	11%	1	7%
Race				
White	13	72%	13	87%
Black or African American	2	11%	1	7%
Hispanic or Latino	2	11%	1	7%
Asian or Asian American	0	0%	0	0%
Native Hawaiian or Other Pacific Islander	0	0%	0	0%
Other	1	6%	0	6%
Gender				
Female	12	67%	10	67%
Male	6	33%	5	33%
Are you currently supervising or mentoring beginning teachers?				
Yes	16	89%	13	87%
No	2	11%	2	13%
Are you currently involved in the training/preparation of teacher candidates?				
Yes	18	100%	15	100%
In all, how many years have you taught teacher candidates?				
3 years or less	2	11%	2	13%
4 - 7 years	5	28%	3	20%
8 - 11 years	3	17%	4	27%
12 - 15 years	1	6%	3	20%
16 years or more	7	39%	3	20%

Table D2

Passing Score Summary by Round of Judgments — Reading

	Panel 1		Panel 2	
Panelist	**Round 1**	**Round 2**	**Round 1**	**Round 2**
1	25.70	26.10	29.70	29.70
2	28.50	28.30	35.70	35.80
3	36.70	37.35	22.00	23.10
4	29.40	29.30	25.85	25.35
5	30.40	31.30	22.00	22.10
6	21.05	29.70	30.10	31.10
7	32.20	32.80	29.70	29.90
8	30.10	30.90	30.00	30.60
9	31.20	30.90	26.65	28.00
10	33.30	33.50	34.85	34.85
11	34.10	34.35	24.90	25.20
12	29.05	28.55	30.55	30.65
13	32.30	32.30	34.20	32.20
14	29.15	30.65	30.70	30.70
15	35.95	35.75	31.95	31.45
16	31.00	31.60	37.55	36.20
17	33.50	33.25	37.20	37.00
18	22.90	22.50		
19	28.95	30.55		
20	27.60	28.00		
21	25.05	25.55		
22	35.35	35.65		
23	28.20	27.50		
24	27.80	28.60		
Average	29.98	30.62	30.21	30.23
Lowest	21.05	22.50	22.00	22.10
Highest	36.70	37.35	37.55	37.00
SD	3.92	3.49	4.80	4.42
SEJ	0.80	0.71	1.16	1.07

Table D3

Passing Score Summary by Round of Judgments — Writing

	Panel 1		Panel 3	
Panelist	**Round 1**	**Round 2**	**Round 1**	**Round 2**
1	42.30	42.30	35.90	46.80
2	42.20	39.90	42.45	42.05
3	51.05	48.75	34.00	34.20
4	42.80	42.40	42.30	42.30
5	40.80	40.00	36.50	37.20
6	43.10	44.50	40.60	41.10
7	43.20	45.60	41.80	41.20
8	38.00	37.80	43.10	42.30
9	45.50	45.20	33.20	46.30
10	45.00	43.00	45.65	43.95
11	37.20	40.10	48.35	48.45
12	42.45	42.25	42.00	44.30
13	42.10	42.10	54.80	54.80
14	40.75	40.25	46.80	46.80
15	48.35	48.15	42.40	42.40
16	44.40	45.50	47.50	47.20
17	44.75	44.75	43.70	43.70
18	41.30	41.20	41.60	42.00
19	43.80	44.00		
20	40.40	42.20		
21	43.10	43.10		
22	48.10	48.10		
23	40.80	40.90		
24	38.90	39.60		
Average	42.93	42.99	42.37	43.73
Lowest	37.20	37.80	33.20	34.20
Highest	51.05	48.75	54.80	54.80
SD	3.21	2.90	5.34	4.48
SEJ	0.66	0.59	1.26	1.06

Table D4

Passing Score Summary by Round of Judgments — Mathematics

	Panel 1		Panel 4	
Panelist	**Round 1**	**Round 2**	**Round 1**	**Round 2**
1	29.00	29.00	27.50	28.00
2	22.80	23.80	35.50	32.95
3	30.90	30.80	24.60	25.90
4	30.70	29.30	24.25	25.85
5	20.50	21.80	34.30	34.15
6	25.80	27.10	26.85	27.65
7	31.80	31.20	28.65	28.50
8	27.70	27.50	23.90	25.60
9	26.60	25.80	33.20	32.70
10	25.75	26.45	34.95	34.30
11	29.40	29.50	29.60	29.50
12	21.85	24.20	28.05	28.75
13	27.60	27.60	32.10	29.65
14	22.60	24.00	33.30	30.70
15	32.90	32.20		
16	28.90	29.20		
17	26.20	24.75		
18	17.40	19.05		
19	33.70	30.70		
20	24.85	25.65		
21	24.90	25.00		
22	29.10	29.60		
23	24.60	24.70		
24	26.50	27.20		
Average	26.75	26.92	29.77	29.59
Lowest	17.40	19.05	23.90	25.60
Highest	33.70	32.20	35.50	34.30
SD	3.99	3.20	4.11	3.00
SEJ	0.81	0.65	1.10	0.80

Table D5

Final Evaluation: Panel 1

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	24	100%	0	0%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	23	96%	1	4%	0	0%	0	0%
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	24	100%	0	0%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	18	75%	6	25%	0	0%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	22	92%	2	8%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	21	88%	3	13%	0	0%	0	0%

Table D5 (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 target candidate	22	92%	2	8%	0	0%
• The between-round discussions	16	67%	8	33%	0	0%
• The knowledge/skills required to answer each test item	19	79%	5	21%	0	0%
• The passing scores of other panel members	6	25%	15	63%	3	13%
• My own professional experience	17	71%	7	29%	0	0%

Overall, how comfortable are you with the panel's recommended passing scores?	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	Percent	*N*	Percent	*N*	Percent	*N*	Percent
• Reading	19	79%	5	21%	0	0%	0	0%
• Writing	18	75%	6	25%	0	0%	0	0%
• Mathematics	17	71%	5	21%	1	4%	1	4%

Overall, the recommended passing score is:	Too Low		About Right		Too High	
	N	Percent	*N*	Percent	*N*	Percent
• Reading	0	0%	24	100%	0	0%
• Writing	0	0%	23	96%	1	4%
• Mathematics	3	13%	21	88%	0	0%

Table D6

Final Evaluation: Panel 2

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	13	76%	4	24%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	14	82%	3	18%	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	65%	6	35%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	12	71%	5	29%	0	0%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	12	71%	5	29%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	9	53%	8	47%	0	0%	0	0%

Table D6 (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 target candidate	10	59%	7	41%	0	0%
• The between-round discussions	9	53%	8	47%	0	0%
• The knowledge/skills required to answer each test item	12	71%	5	29%	0	0%
• The passing scores of other panel members	1	6%	13	76%	3	18%
• My own professional experience	14	82%	3	18%	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?	10	59%	6	35%	1	6%	0	0%

	Too Low		About Right		Too High	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0%	16	94%	1	6%

Table D7

Final Evaluation: Panel 3

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	16	89%	2	11%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	16	89%	2	11%	0	0%	0	0%
• The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	13	72%	5	28%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	11	61%	6	33%	1	6%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	14	78%	3	17%	1	6%	0	0%
• The process of making the standard-setting judgments was easy to follow.	12	67%	6	33%	0	0%	0	0%

Table D7 (continued)

Final Evaluation: Panel 3

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 target candidate	7	39%	11	61%	0	0%
• The between-round discussions	8	44%	9	50%	1	6%
• The knowledge/skills required to answer each test item	13	72%	5	28%	0	0%
• The passing scores of other panel members	2	11%	11	61%	5	28%
• My own professional experience	13	72%	5	28%	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?	7	39%	7	39%	3	17%	1	6%

	Too Low		About Right		Too High	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0%	15	83%	3	17%

Table D8

Final Evaluation: Panel 4

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	**%**	***N***	**%**	***N***	**%**	***N***	**%**
• I understood the purpose of this study.	15	100%	0	0%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	15	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.	15	100%	0	0%	0	0%	0	0%
• The explanation of how the recommended passing score is computed was clear.	10	67%	4	27%	1	7%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	12	80%	3	20%	0	0%	0	0%
• The process of making the standard-setting judgments was easy to follow.	12	80%	3	20%	0	0%	0	0%

Table D8 (continued)

Final Evaluation: Panel 4

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 target candidate	10	67%	5	33%	0	0%
• The between-round discussions	10	67%	5	33%	0	0%
• The knowledge/skills required to answer each test item	9	60%	6	40%	0	0%
• The passing scores of other panel members	4	27%	9	60%	2	13%
• My own professional experience	8	53%	7	47%	0	0%

	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	%	*N*	%	*N*	%	*N*	%
• Overall, how comfortable are you with the panel's recommended passing score?	13	87%	2	13%	0	0%	0	0%

	Too Low		About Right		Too High	
	N	%	*N*	%	*N*	%
• Overall, the recommended passing score is:	0	0%	15	100%	0	0%