# Multistate Standard-Setting Technical Report

# *PRAXIS*® MATHEMATICS (5165)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2021

Copyright © 2021 by ETS. All rights reserved. ETS, the ETS logo, and PRAXIS are registered trademarks of ETS. 44905976

# 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 |

Listening. Learning. Leading.

# Multi-State Standard Setting Technical Report

# PRAXIS ART: CONTENT KNOWLEDGE (0134)

Educational and Credentialing Research

Educational Testing Service

Princeton, New Jersey

November 2010

Copyright © 2010 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.

## Executive Summary

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for a revised assessment in the Praxis Series™ — Art: Content Knowledge (0134) — research staff from Educational Testing Service (ETS) designed and conducted two multi-state standard setting studies[1]. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level Art teachers.

### Participating States

Panelists from 22 states were recommended by state departments of education to participate on expert panels. The state departments of education recommended panelists with (a) art education experience, either as K-12 Art teachers or college faculty who prepare Art teachers and (b) familiarity with the knowledge and skills required of beginning Art teachers.

### Recommended Cut Scores

The recommended cut scores for each panel, as well as the average cut score across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score. For the Praxis Art: Content Knowledge (0134) assessment, the average recommended cut score (rounded up) is 72 (on the raw score metric), which represents 65% of total available 110 raw score points (the recommended cut scores for Panels 1 and 2 are 73 and 71, respectively). The scaled score associated with a raw score of 72 is 158.

### Summary of Content Specification Judgments

Panelists judged the extent to which the knowledge and/or skills reflected by the content specifications were important for entry-level Art teachers. The favorable judgments of the panelists provided evidence that the content of the assessment is important for beginning practice.

[1] The two multi-state standard setting studies collected expert judgments for questions comprising both the Praxis Art: Content Knowledge (0134) and Praxis Art: Content and Analysis (0135) assessments. Standard-setting procedures and results presented in the following report only pertain to the Praxis Art: Content Knowledge (0134) assessment. A separate report contains similar information for Praxis Art: Content and Analysis (0135).

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for a revised assessment in the Praxis Series™ — Art: Content Knowledge (0134) — research staff from Educational Testing Service (ETS) designed and conducted two multi-state standard setting studies[2]. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level Art teachers. Panelists were recommended by state departments of education[3] to participate on the two expert panels. The state departments of education recommended panelists with (a) art education experience, either as K-12 Art teachers or college faculty who prepare Art teachers and (b) familiarity with the knowledge and skills required of beginning Art teachers.

The two, non-overlapping panels (a) allow each participating state to be represented and (b) provide a replication of the judgment process to strengthen the technical quality of the recommended passing score. For the Praxis Art: Content Knowledge (0134) assessment, 22 states were represented by 42 panelists across the two panels, (See Appendix A for the names and affiliations of the panelists.)

[2] The two multi-state standard setting studies collected expert judgments for questions comprising both the Praxis Art: Content Knowledge (0134) and Praxis Art: Content and Analysis (0135) assessments. Standard-setting procedures and results presented in the following report only pertain to the Praxis Art: Content Knowledge (0134) assessment. A separate report contains similar information for Praxis Art: Content and Analysis (0135).

[3] State departments of education that currently use one or more Praxis tests were invited to participate in the multi-state standard setting studies.

**Table 1**

**Participating States (and number of panelists) for Multi-State Panels**

| | |
|---|---|
| Alabama (2 panelists) | New Hampshire (2 panelists) |
| Arkansas (2 panelists) | New Jersey (2 panelists) |
| Connecticut (2 panelists) | Ohio (2 panelists) |
| Kentucky (2 panelists) | Pennsylvania (1 panelist) |
| Louisiana (2 panelists) | South Carolina (2 panelists) |
| Maryland (2 panelists) | Tennessee (2 panelists) |
| Maine (2 panelists) | Utah (2 panelists) |
| Missouri (2 panelists) | Vermont (2 panelists) |
| Mississippi (2 panelists) | Washington, DC (1 panelist) |
| North Carolina (2 panelists) | Wisconsin (2 panelists) |
| North Dakota (2 panelists) | West Virginia (2 panelists) |

*Note*. Pennsylvania and Washington, DC were represented on only one of the two panels.

The panels were convened in November 2010 in Princeton, New Jersey. For both panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing score, or cut score.

The following technical report is divided into three sections. The first section describes the content and format of the assessment. The second section describes the standard setting processes and methods used. The third section presents the results of the standard setting studies.

The passing score recommendation for the assessment is provided to each of the represented state departments of education. In each state, the department of education, the state board of education, or a designated educator licensure board is responsible for establishing the final passing score in accordance with applicable state regulations.

The first national administration of the new Praxis Art: Content Knowledge (0134) assessment will occur in fall 2011.

## Overview of the Praxis Assessment

The *Praxis Art: Content Knowledge Test at a Glance* document (ETS, in press) describes the purpose and structure of the assessment. In brief, the assessment measures whether entry-level Art teachers have the content knowledge believed necessary for competent professional practice. A National Advisory Committee of expert practitioners and preparation faculty defined the content of the assessment, and a national survey of the field confirmed the content.

The two hour assessment contains 120 multiple-choice questions covering *Art Making- General* (approximately 15 questions); *Art Making – Media & Processes* (approximately 61 questions); *Materials & Processes in a Historical Context and Responding to Art* (approximately 17 questions); and *Western Tradition and Beyond the Western Tradition* (approximately 27 questions)[4].

Candidate scores are reported as an overall score; four category scores – one for each content area listed above – also are reported. Of the 120 multiple-choice questions, 110 questions contribute to a candidate's score[5]. The maximum total number of raw points that may be earned on each assessment is 110. The reporting scale for the Praxis Art: Content Knowledge (0134) assessment ranges from 100 to 200 scaled-score points.

## Processes and Methods

For both expert panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing score, or cut score. The following section describes the processes and methods used[6]. (The agenda for the panel meetings is presented in Appendix A.)

The design of the standard setting study included two non-overlapping expert panels. The training provided to panelists as well as the study materials were consistent across panels with the exception of defining the Just Qualified Candidate (JQC). To assure that both panels were using the same frame of reference when making question-level standard setting judgments, the JQC definition developed through a consensus process by the first panel was used as the definition for the second panel.

[4] The number of questions for each content area may vary slightly from form to form of the assessment.
[5] Ten of the 120 multiple-choice questions are pretest questions which do not contribute to a candidate's score.
[6] Panelists also judged the constructed-response questions that appear on the Praxis Art: Content and Analysis assessment. The process for making theses judgments are not described in this report but are described in the technical report for the Praxis Art: Content and Analysis (0135) standard setting.

The second panel did complete a thorough review of the definition to allow panelists to internalize the definition. The processes for developing the definition (with Panel 1) and reviewing/internalizing the definition (with Panel 2) are described later, and the Just Qualified Candidate definitions are presented in Appendix C.

The panelists were sent an e-mail explaining the purpose of the standard-setting study and requesting that they review the test content specifications for the assessment (included in the *Test at a Glance* document, which was attached to the e-mail). The purpose of the review was to familiarize the panelists with the general structure and content of the assessment.

The standard-setting study began with a welcome and introduction by the meeting facilitator, Dr. Clyde Reese from the Center for Validity Research. He explained how the assessment was developed, provided an overview of standard setting, and presented the agenda for the study.

**Reviewing the Assessment**

The first activity was for the panelists to "take the test." (Each panelist had signed a nondisclosure form.) The panelists were given approximately an hour and a half to respond to the 120 multiple-choice questions (as well as the three constructed-response questions that are included on the Praxis Art: Knowledge and Analysis assessment). Panelists were instructed not to refer to the answer key while taking the test. The purpose of "taking the test" was for the panelists to become familiar with the test format, content, and difficulty. After "taking the test," the panelists checked their responses against the answer key.

The panelists then engaged in a discussion of the major content areas being addressed by the assessment; they were also asked to remark on any content areas that they thought would be particularly challenging for entering Art teachers, and areas that addressed content that would be particularly important for entering teachers.

**Defining the Just Qualified Candidate**

Following the review of the assessment, panelists internalized the definition of the Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of knowledge and/or skills believed necessary to be a qualified Art teacher. The JQC definition is the operational definition of the

cut score. The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC.

In Panel 1, the panelists were split into smaller groups, and each group was asked to write down their definition of a JQC. Each group referred to the *Praxis Art: Content and Analysis Test at a Glance*[7] to guide their definition. Each group posted its definition on chart paper, and a full-panel discussion occurred to reach consensus on a definition (see Appendix C).

In Panel 2, the panelists began with the definition of the JQC developed by the first panel. Given that the multi-state standard setting study was designed to replicate processes and procedures across the two panels, it was important that both panels use consistent JQC definitions to frame their judgments. For Panel 2, the panelists reviewed the JQC definition, and any ambiguities were discussed and clarified. The panelists then were split into smaller groups, and each group developed performance indicators or "can do" statements based on the definition. The purpose of the indicators was to provide clear examples of what might be observed to indicate that the teacher had the defined knowledge. The performance indicators were shared across the group, and discussed and added to the definition. The panel also had an opportunity to suggest minor changes to the initial definition, if doing so added clarity. Panel 2 made two revisions to the JQC definition developed by Panel 1: (a) splitting the first bullet into two separate bullets and (b) replacing one of the examples in the second bullet.

**Panelists' Judgments**

A probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) was used for the multiple-choice questions. In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC would answer it 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 a JQC would answer the question correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly.

For each panel, the panelists were asked to approach the judgment process in two stages. First, they reviewed the definition of the JQC and the question and decided if, overall, the question was

[7] The test specifications contained in the *Praxis Art: Content and Analysis Test At A Glance* subsumed the specification for the Praxis Art: Content Knowledge (0134) assessment.

difficult for the JQC, easy for the JQC, or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rule of thumb to guide their decision:

- difficult questions for a JQC were in the 0 to .30 range;
- moderately difficult/easy questions for a JQC were in the .40 to .60 range; and
- easy questions for a JQC were in the .70 to 1 range.

The second decision was for panelists to decide how they wanted to refine their judgment within the range. For example, if a panelist thought that a question was easy for a JQC, the initial decision located the question in the .70 to 1 range. The second decision was for the panelist to decide if the likelihood of answering it correctly was .70, .80, .90, .95, or 1.0. The two-stage decision-process was implemented to reduce the cognitive load placed on the panelists. The panelists practiced making their standard-setting judgments on six of the multiple-choice questions.

The panelists engaged in two rounds of judgments. Following Round 1, question-level feedback was provided to the panel. The panelists' judgments were displayed for each question. The panelists' judgments were summarized by the three general difficulty levels (0 to .30, .40 to .60, and .70 to 1), and the panel's average question judgment was provided. Questions were highlighted to show when panelists converged in their judgments (at least two-thirds of the panelists located a question in the same difficulty range) or diverged in their judgments. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2).

Other than the definition of the JQC, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

**Judgment of Content Specifications**

In addition to the two-round standard setting process, each panel judged the importance of the knowledge and/or skills stated or implied in the assessment content specifications for the job of an entry-level Art teacher. These judgments addressed the perceived content-based validity of the assessment. Judgments were made using a four-point Likert scale — V*ery Important*, I*mportant*, S*lightly Important*,

and N*ot Important*. Each panelist independently judged the knowledge categories and knowledge/skills statements.

## Results

Results are presented separately for the two panels. The recommended cut scores for each panel, as well as the average cut score across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score.

### Expert Panels

The standard setting study included two expert panels. The various state departments of education recruited panelists to represent a range of professional perspectives. A description of the panels is presented below. (See Appendix A for a listing of panelists for each panel.)

Panel 1 included 21 teachers, administrators, and college faculty who prepare Art teachers, representing 21 states. In brief, 18 panelists were teachers, one was an administrators or department heads, and two were college faculty. Both of the panelists who were college faculty were currently involved in the training or preparation of Art teachers. Seventeen panelists were White, one was Hispanic or Latino, one was Asian American, one was American Indian or Alaskan Native, and one panelist indicated "other." Fifteen panelists were female. Nineteen panelists reported being certified Art teachers in their states. Slightly less than half of panelists (9 of the 21 panelists or 43%) had seven or fewer years of experience as an Art teacher, and five had 16 or more years of teaching experience.

Panel 2 included 21 teachers and college faculty who prepare Art teachers, representing 21 states. In brief, 18 panelists were teachers and three were college faculty. All three of the panelists who were college faculty were currently involved in the training or preparation of Art teachers. Seventeen panelists were White, three were African American, and one was Asian American. Fourteen panelists were female. Eighteen panelists reported being certified Art teachers in their states. Nearly 40% of panelists (8 of the 21 panelists) had seven or fewer years of experience as an Art teacher, and three had 16 or more years of teaching experience.

A fuller demographic description for the members of the two panels is presented in Tables D1 and D2 in Appendix D.

**Initial Evaluation Forms.**

The panelists completed an initial evaluation after receiving training on how to make question-level judgments. The primary information collected from this form was the panelists indicating if they had received adequate training to make their standard-setting judgments and were ready to proceed. Across both panels, all panelists indicated that they were prepared to make their judgments.

**Summary of Standard Setting Judgments by Round.**

A summary of each round of standard-setting judgments is presented in Appendix D. The numbers in each table reflect the recommended cut scores — the number of raw points needed to "pass" the assessment — of each panelist for the two rounds. The panel's average recommended cut score and highest and lowest cut scores are reported, as are the standard deviations (SD) of panelists' cut scores and the standard errors of judgment (SEJ). The SEJ is one way of estimating the reliability of the judgments. It indicates how likely it would be for other panels of educators similar in make-up, experience, and standard-setting training to the current panels to recommend the same cut score on the same form of the assessment. A comparable panel's cut score would be within 1 SEJ of the current average cut score 68 percent of the time and within 2 SEJs 95 percent of the time.

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 both panels. The Round 2 average total score is the panel's recommended cut score (passing score).

The panels' cut score recommendations for the Praxis Art: Content Knowledge (0134) assessment are 72.79 for Panel 1 and 70.33 for Panel 2 (see Tables D3 and D4 in Appendix D). The values were rounded to the next highest whole number to determine the functional recommended cut scores — 73 for Panel 1 and 71 for Panel 2. The values of 73 and 71 represent approximately 66% and 65%, respectively, of the total available 110 raw-score points that could be earned on the assessment. The scaled scores associated with 73 and 71 raw points are 159 and 156, respectively.[8]

[8] For reference purposes, if the recommended raw cut score was 72 or 70 points, the scaled score would be 158 or 155, respectively.

Table D5 (in Appendix D) present the estimated standard errors of measurement (SEM) around the recommended cut scores for each panel. A standard error represents the uncertainty associated with a test score. The scaled scores associated with 1 and 2 SEMs above and below the recommended cut scores are provided. The standard errors provided are an estimate, given that the Praxis Art: Content Knowledge (0134) assessment has not yet been administered.

In addition to the recommended cut score for each panel, the average cut score across the two panels is provided to help state departments of education determine an appropriate cut (or passing) score for the Praxis Art: Content Knowledge (0134) assessment. The panels' average cut score recommendation for the Praxis Art: Content Knowledge (0134) assessment is 71.56. The value was rounded to 72 (next highest raw score) to determine the functional recommended cut score. The value of 72 represents approximately 65% of the total available 110 raw-score points that could be earned on the assessment. The scaled score associated with 72 raw points is 158. Table D5 (in Appendix D) presents the standard error of measurement (SEM) around the recommended cut score combining the information from the two panels.

**Summary of Content Specification Judgments.**

Panelists judged the extent to which the knowledge and/or skills reflected by the Praxis Art: Content Knowledge (0134) assessment content specifications were important for entry-level Art teachers. Panelists rated the knowledge categories and knowledge/skills statements on a four-point scale ranging from *Very Important* to *Not Important*. The panelists' ratings are summarized in Table D6 (in Appendix D).

Thirteen of the 16 knowledge categories were judged to be *Very Important* or *Important* by 90% or more of the panelists. The knowledge categories of *"Understanding and Applying the Elements of Art and Principles of Visual Organization,"* and *"Knows and Understands Safety, Environment, and Storage Issue"* (81% of the panelists judged as *Very Important*) were seen as the most important for beginning Art teachers. The knowledge category of *"Understanding Materials, Tools and Processes for Videography, Filmmaking, and Installations"* (65% of the panelists judged as *Slightly Important* or *Not Important*) was seen as less important for beginning Art teachers. All but four of the 70 knowledge statements covered by the Praxis Art: Content Knowledge assessment were judged to be *Very Important* or *Important* by at least two-thirds of the panelists.

**Summary of Final Evaluations**

The panelists completed an evaluation form at the conclusion of their standard setting study. The evaluation form asked the panelists to provide feedback about the quality of the standard-setting implementation and the factors that influenced their decisions. Tables D7 and D8 (in Appendix D) present the results of the final evaluations.

All panelists *agreed* or *strongly agreed* that they understood the purpose of the study and that the facilitator's instructions and explanations were clear. All panelists *agreed* or *strongly agreed* that they were prepared to make their standard setting judgments. Across the two panels, all but one of the panelists *strongly agreed* or *agreed* that the standard-setting process was easy to follow.

All panelists reported that the definition of the JQC was at least *somewhat influential* in guiding their standard-setting judgments; 86% of panelists indicated the definition was *very influential*. All but two of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Nearly three-quarters of the panelists (32 of the 42 panelists) indicated that the knowledge/skills required to answer each question as *very influential* in guiding their judgments.

There were similar ratings between the two panels when asked to respond to their level of comfort with their panel's recommended passing score[9]. All panelists indicated they were *very* or *somewhat comfortable* with their recommendation. For both panels, the majority of the panelists indicated that the recommend cut score was *about right* (100% for Panel 1 and 90% for Panel 2). Of the remaining panelists from Panel 2, one indicated the cut score was *too low* and one indicated it was *too high*.

[9] Panelists indicated their level of comfort with the cut score recommendations for both the Praxis Art: Content Knowledge (0134) and Praxis Art: Content and Analysis (0135) assessments.

## Summary

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for a revised assessment in the Praxis Series™ — Art: Content Knowledge (0134) — research staff from Educational Testing Service (ETS) designed and conducted two multi-state standard setting studies[10]. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level Art teachers.

The recommended cut scores for each panel, as well as the average cut score across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score. For the Praxis Art: Content Knowledge (0134) assessment, the average recommended cut score (rounded up) is 72 (on the raw score metric), which represents 65% of total available 110 raw score points (the recommended cut scores for Panels 1 and 2 are 73 and 71, respectively). The scaled score associated with a raw score of 72 is 158.

Panelists judged the extent to which the knowledge and/or skills reflected by the content specifications were important for entry-level Art teachers. The favorable judgments of the panelists provided evidence that the content of the assessment is important for beginning practice.

[10] The two multi-state standard setting studies collected expert judgments for questions comprising both the Praxis Art: Content Knowledge (0134) and Praxis Art: Content and Analysis (0135) assessments. Standard-setting procedures and results presented in the following report only pertain to the Praxis Art: Content Knowledge (0134) assessment. A separate report contains similar information for Praxis Art: Content and Analysis (0135).

## References

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

ETS. (In press). *Art: Content Knowledge Test at a Glance*. Princeton, NJ.

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.

## Appendix A

## Panelists' Names & Affiliations

**Praxis Art: Content Knowledge & Praxis Art: Content and Analysis**

**Panel 1**

| Panelist | Affiliation |
|---|---|
| Bonner, Bethany | Oakdale Elementary School (CT) |
| Brasser, Angela | Campbellsville Middle & High Schools (KY) |
| Brouillette, Charles | A. E. Phillips Lab School\Louisiana Tech University (LA) |
| Coon, John Mark | Canton High School (MS) |
| Cowles, Mariam | Cedarville School District (AR) |
| Cullinan, Mary Susan | Colliers Primary (WV) |
| Gall, Marta | Macon R-1 School (MO) |
| Heid, Karen A. | University of South Carolina (SC) |
| Hernández-Balcázar, Noemí Verónica | Kearns High School (UT) |
| LeCours, Elizabeth | Hardwick Elementary School (VT) |
| Lindsey, Jennifer | Mooresville Intermediate School (NC) |
| Milliken, Chris | Wells Junior High School (ME) |
| Mock, Stephen | Memphis City Schools (TN) |
| Mojzsis, Katherine | Sayreville War Memorial High School (NJ) |
| Morin, Derek | Kindred Public School (ND) |
| Northcutt, Adriana E. | Trace Crossings School (AL) |
| O'Gorman Rhodebeck, Kathleen | Pembroke Hill School (NH) |
| Purcell Sacco, Kristine | ACLD Tillotson School (PA) |
| Roemer, Jordyn M. | North County High School (MD) |
| Schorsch, Jamie | Oak Hills High School (OH) |
| Tarrell, Robert | Edgewood College (WI) |

**Praxis Art: Content Knowledge & Praxis Art: Content and Analysis (continued)**

**Panel 2**

| **Panelist** | **Affiliation** |
|---|---|
| Aman, Ronald | West Virginia University (WV) |
| Armstead, Jacqueline | Argyle Middle School (MD) |
| Csejtey, Stephen | Akron Public Schools (OH) |
| Danenhauer, Audrea | Farmington Public Schools (AR) |
| Dieck, Jessica M. | Hinds County School District (MS) |
| Dunn, Holli J. | Kickapoo High School (MO) |
| Edinger, Ted | Tulip Grove Elementary MNPS (TN) |
| England, Marla | Barren County Middle School (KY) |
| Foley, Lisa | Chittenden Central SU (VT) |
| Gumbulevich, Jeanette | Waterbury Arts Magnet School (CT) |
| Hill, Bryan | MLKing Elementary School (DC) |
| Kerrigan, Danette | Sacopee Valley Middle School (ME) |
| Leach, Randall | J.H. Rose High School (NC) |
| Parsons, Juliella | Tuscaloosa City School System (AL) |
| Roberts, Kathryn | Spanish Fork High School (UT) |
| Skow, Margaret | Rollings Middle School of the Arts\Dorchester District Two (SC) |
| Summers, Bridget | Lakewood Elementary School (LA) |
| Swift, Jason | Plymouth State University (NH) |
| Wilkie, Kenneth | Riverside School, Princeton (NJ) |
| Winker, Melissa | Memorial High School (WI) |
| Yang, Crystal | University of North Dakota (ND) |

## Appendix B

## Workshop Agenda

**Praxis Art: Content Knowledge and**

**Praxis Art: Content and Analysis Assessments**

**Standard Setting Study**

**Day 1**

| | |
|---|---|
| **8:00 – 8:15** | **Welcome and Introduction**<br>* **Overview of Workshop Events** |
| **8:15 – 8:45** | **Overview of Standard Setting & the Praxis Art Assessments** |
| **8:45 – 9:00** | **Break** |
| **9:00 – 10:30** | **"Take" the Praxis Art Assessments** |
| **10:30 – 11:00** | **Discuss the Praxis Art Assessments** |
| **11:00 – 12:00** | **Define the Knowledge/Skills of a JQC** |
| **12:00 – 12:45** | **Lunch** |
| **12:45 – 2:15** | **Define the Knowledge/Skills of a JQC (continued)** |
| **2:15 – 2:30** | **Break** |
| **2:30 – 3:00** | **Standard Setting Training for CR Questions** |
| **3:00 – 3:30** | **Round 1 Standard Setting Judgments: CR Questions** |
| **3:30 – 4:00** | **Standard Setting Training for MC Questions** |
| **4:00 – 5:00** | **Round 1 Standard Setting Judgments: MC Questions 1- 40** |
| **5:00 – 5:15** | **Collect Materials; End of Day 1** |

**Praxis Art: Content Knowledge and**

**Praxis Art: Content and Analysis Assessments**

**Standard Setting Study**

**Day 2**

| | |
|---|---|
| **9:00 – 9:05** | **Overview of Day 2** |
| **9:05 – 9:15** | **Review Standard Setting for MC Questions** |
| **9:15 – 10:30** | **Round 1 Standard Setting Judgments: MC Questions 41- 120** |
| **10:30 – 10:45** | **Break** |
| **10:45 – 11:15** | **Round 1 Feedback & Round 2 Judgments: CR Questions** |
| **11:15 – 12:00** | **Round 1 Feedback & Round 2 Judgments: MC Questions** |
| **12:00 – 1:00** | **Lunch** |
| **1:00 – 2:15** | **Round 1 Feedback & Round 2 Judgments: MC Questions (continued)** |
| **2:15 – 3:00** | **Specification Judgments** |
| **3:00 – 3:15** | **Feedback on Round 2 Recommended Cut Score** |
| **3:15 – 3:30** | **Complete Final Evaluation** |
| **3:30 – 3:45** | **Collect Materials; End of Study** |

# Appendix C

## Just Qualified Candidate (JQC) Definitions

**Description of a Just Qualified Candidate**
**Panel 1**

**A JQC …**

1. Knows characteristics of common 2-D and 3-D materials and processes and compare across materials and processes

2. Knows characteristics of common technologies and equipment such as printmaking, photography, film making, and computers

3. Understands safety and health issues related to common materials and processes; applies safety procedures in the classroom

4. Can prepare an exhibition demonstrating an understanding of aesthetic presentation

5. Demonstrates knowledge and application of art vocabulary

6. Knows major trends in Western and Nonwestern art and architecture

7. Knows and understands the chronological timeline and thematic organization of art history

8. Ability to analyze works of art and evaluate them critically across cultures and periods of time

9. Understands the roles of function and purpose of works of art (i.e., reflection)

10. Knows the role of visual literacy and popular culture

**Description of a Just Qualified Candidate**
**Panel 2**

**A JQC …**

1. Knows characteristics of common 2-D and 3-D materials and processes
2. Can compare characteristics of common 2-D and 3-D materials and processes.
3. Knows characteristics of common technologies and equipment such as printmaking, photography, videography, and computer applications
4. Understands safety and health issues related to common materials and processes; applies safety procedures in the classroom
5. Can prepare an exhibition demonstrating an understanding of aesthetic presentation
6. Demonstrates knowledge and application of art vocabulary
7. Knows major trends in Western and Nonwestern art and architecture
8. Knows and understands the chronological timeline and thematic organization of art history
9. Ability to analyze works of art and evaluate them critically across cultures and periods of time
10. Understands the roles of function and purpose of works of art (i.e., reflection)
11. Knows the role of visual literacy and popular culture

## Appendix D

# Results for Praxis Art: Content Knowledge

**Table D1**

**Panel Member Demographics — Panel 1**

| | N | Percent |
|---|---|---|
| **Current Position** | | |
| Teachers | 18 | 86% |
| Teacher/Administrator | 1 | 5% |
| College Faculty | 2 | 10% |
| **Race** | | |
| White | 17 | 81% |
| Hispanic or Latino | 1 | 5% |
| Asian or Asian American | 1 | 5% |
| American Indian or Alaskan Native | 1 | 5% |
| Other | 1 | 5% |
| **Gender** | | |
| Female | 15 | 71% |
| Male | 6 | 29% |
| **Are you currently certified as an Art teacher in your state?** | | |
| Yes | 19 | 90% |
| No | 2 | 10% |
| **Are you currently teaching Art in your state?** | | |
| Yes | 21 | 100% |
| No | 0 | 0% |
| **Are you currently supervising or mentoring other Art teachers?** | | |
| Yes | 9 | 43% |
| No | 12 | 57% |
| **How many years of experience do you have teaching Art?** | | |
| 3 years or less | 1 | 5% |
| 4 - 7 years | 8 | 38% |
| 8 - 11 years | 5 | 24% |
| 12 - 15 years | 2 | 10% |
| 16 years or more | 5 | 24% |

**Table D1 (continued)**

**Panel Member Demographics — Panel 1**

| | N | Percent |
|---|---|---|
| **At what K-12 grade level are you currently teaching Art?** | | |
| Elementary (K - 5 or K - 6) | 8 | 38% |
| Middle School (6 - 8 or 7 - 9) | 1 | 5% |
| Elementary and Middle School | 2 | 10% |
| High School (9 - 12 or 10 - 12) | 5 | 24% |
| Middle and High School | 2 | 10% |
| All Grades | 1 | 5% |
| Not currently teaching at the K-12 level | 2 | 10% |
| **Which best describes the location of your K-12 school?** | | |
| Urban | 3 | 14% |
| Suburban | 10 | 48% |
| Rural | 6 | 29% |
| Not currently working in a K-12 school | 2 | 10% |
| **If you are college faculty, are you currently involved in the training/preparation of Art teachers?** | | |
| Yes | 2 | 10% |
| No | 0 | 0% |
| Not college faculty | 19 | 90% |

**Table D2**

**Panel Member Demographics — Panel 2**

| | N | Percent |
|---|---|---|
| **Current Position** | | |
| Teachers | 18 | 86% |
| College Faculty | 3 | 14% |
| **Race** | | |
| White | 17 | 81% |
| Black or African American | 3 | 14% |
| Asian or Asian American | 1 | 5% |
| **Gender** | | |
| Female | 14 | 67% |
| Male | 7 | 33% |
| **Are you currently certified as an Art teacher in your state?** | | |
| Yes | 18 | 86% |
| No | 3 | 14% |
| **Are you currently teaching Art in your state?** | | |
| Yes | 21 | 100% |
| No | 0 | 0% |
| **Are you currently supervising or mentoring other Art teachers?** | | |
| Yes | 10 | 48% |
| No | 11 | 52% |
| **How many years of experience do you have teaching Art?** | | |
| 3 years or less | 2 | 10% |
| 4 - 7 years | 6 | 29% |
| 8 - 11 years | 6 | 29% |
| 12 - 15 years | 4 | 19% |
| 16 years or more | 3 | 14% |

**Table D2 (continued)**

**Panel Member Demographics — Panel 2**

| | N | Percent |
|---|---|---|
| **At what K-12 grade level are you currently teaching Art?** | | |
| Elementary (K - 5 or K - 6) | 8 | 38% |
| Middle School (6 - 8 or 7 - 9) | 3 | 14% |
| Elementary and Middle School | 1 | 5% |
| High School (9 - 12 or 10 - 12) | 5 | 24% |
| Middle and High School | 1 | 5% |
| Not currently teaching at the K-12 level | 3 | 14% |
| **Which best describes the location of your K-12 school?** | | |
| Urban | 8 | 38% |
| Suburban | 7 | 33% |
| Rural | 3 | 14% |
| Not currently working in a K-12 school | 3 | 14% |
| **If you are college faculty, are you currently involved in the training/preparation of Art teachers?** | | |
| Yes | 3 | 14% |
| No | 0 | 0% |
| Not college faculty | 18 | 86% |

**Table D3**

**Cut score Summary by Round of Judgments — Panel 1**

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 68.75 | 72.30 |
| 2 | 74.30 | 74.45 |
| 3 | 69.55 | 71.55 |
| 4 | 54.05 | 54.05 |
| 5 | 61.35 | 62.55 |
| 6 | 58.55 | 59.05 |
| 7 | 72.50 | 73.70 |
| 8 | 75.15 | 75.15 |
| 9 | 73.85 | 74.75 |
| 10 | 72.65 | 70.00 |
| 11 | 72.15 | 72.15 |
| 12 | 72.95 | 73.05 |
| 13 | 82.65 | 82.75 |
| 14 | 75.65 | 74.10 |
| 15 | 81.10 | 80.40 |
| 16 | 75.70 | 75.40 |
| 17 | 72.25 | 73.25 |
| 18 | 93.00 | 93.20 |
| 19 | 69.35 | 70.65 |
| 20 | 68.40 | 68.00 |
| 21 | 78.65 | 78.10 |
| **Average** | 72.50 | 72.79 |
| **SD** | 8.30 | 8.08 |
| **SEJ** | 1.81 | 1.76 |
| **Highest** | 93.00 | 93.20 |
| **Lowest** | 54.05 | 54.05 |

**Table D4**

**Cut score Summary by Round of Judgments — Panel 2**

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 73.05 | 72.55 |
| 2 | 68.75 | 71.40 |
| 3 | 68.30 | 70.20 |
| 4 | 84.75 | 84.20 |
| 5 | 61.80 | 66.70 |
| 6 | 59.60 | 62.65 |
| 7 | 75.75 | 75.55 |
| 8 | 78.80 | 78.00 |
| 9 | 53.70 | 56.50 |
| 10 | 76.30 | 74.60 |
| 11 | 67.85 | 71.45 |
| 12 | 55.10 | 56.20 |
| 13 | 66.55 | 66.45 |
| 14 | 68.00 | 67.80 |
| 15 | 66.60 | 67.65 |
| 16 | 75.30 | 79.50 |
| 17 | 61.40 | 61.30 |
| 18 | 79.10 | 77.90 |
| 19 | 69.05 | 69.85 |
| 20 | 63.25 | 66.75 |
| 21 | 80.95 | 79.65 |
| **Average** | 69.24 | 70.33 |
| **SD** | 8.41 | 7.48 |
| **SEJ** | 1.84 | 1.63 |
| **Highest** | 84.75 | 84.20 |
| **Lowest** | 53.70 | 56.20 |

**Table D5**

**Cut Scores within 1 and 2 SEMs of the Recommended Cut Score**

**(a) Panel 1**

| Recommended Cut Score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 73 (4.98) | | 159 |
| - 2 SEMs | 64 | 147 |
| -1 SEM | 69 | 154 |
| +1 SEM | 78 | 166 |
| + 2 SEMs | 83 | 172 |

**(b) Panel 2**

| Recommended Cut Score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 71 (5.04) | | 156 |
| - 2 SEMs | 61 | 144 |
| -1 SEM | 66 | 150 |
| +1 SEM | 77 | 164 |
| + 2 SEMs | 82 | 171 |

**(c) Combined Across Panels**

| Recommended Cut Score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 72 (5.01) | | 158 |
| - 2 SEMs | 62 | 145 |
| -1 SEM | 67 | 151 |
| +1 SEM | 78 | 166 |
| + 2 SEMs | 83 | 172 |

*Note*. Consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.

**Table D6**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| **I. A. ART MAKING - GENERAL** | | | | | | | | |
| • **Understands and applies the elements of art and principles of visual organization as applied to two-dimensional and three-dimensional media[11].** | **34** | **81%** | **7** | **17%** | **0** | **0%** | **0** | **0%** |
| • Identifies elements and principles of design in visual stimuli | 34 | 81% | 8 | 19% | 0 | 0% | 0 | 0% |
| • Explains relationships of elements to principles | 23 | 55% | 17 | 40% | 2 | 5% | 0 | 0% |
| • Distinguishes uses of elements and principles in two-dimensional and three-dimensional art | 26 | 62% | 15 | 36% | 1 | 2% | 0 | 0% |
| • **Knows various historical methods and contemporary approaches to creating art.** | **12** | **29%** | **26** | **62%** | **4** | **10%** | **0** | **0%** |
| • Defines/identifies both historical and contemporary methods | 10 | 24% | 24 | 57% | 8 | 19% | 0 | 0% |

[11] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| **I. B. ART MAKING – MEDIA AND PROCESSES** | | | | | | | | |
| • **Knows and understands safety, environmental, and storage issues related to the use of art materials and art processes.** | **34** | **81%** | **8** | **19%** | **0** | **0%** | **0** | **0%** |
| • Identifies dangerous materials and their effects | 35 | 83% | 7 | 17% | 0 | 0% | 0 | 0% |
| • Categorizes dangerous materials and their effects | 23 | 55% | 15 | 36% | 4 | 10% | 0 | 0% |
| • Describes proper ventilation, storage, and disposal procedures based on the medium | 26 | 62% | 16 | 38% | 0 | 0% | 0 | 0% |
| • Demonstrates knowledge of MSDS sheets | 14 | 33% | 22 | 52% | 5 | 12% | 1 | 2% |
| • Demonstrates understanding of safety procedures and precautions for using artist's materials and tools | 33 | 79% | 9 | 21% | 0 | 0% | 0 | 0% |
| • Demonstrates knowledge of health issues related to the use of artists' materials and tools | 28 | 67% | 12 | 29% | 2 | 5% | 0 | 0% |
| • **Knows and understands how to use a variety of drawing, painting, and printmaking materials and processes[12].** | **25** | **60%** | **15** | **36%** | **0** | **0%** | **0** | **0%** |
| • Identifies characteristics of materials | 26 | 62% | 16 | 38% | 0 | 0% | 0 | 0% |
| • Identifies similarities and differences among materials | 18 | 43% | 21 | 50% | 3 | 7% | 0 | 0% |
| • Knows vocabulary related to drawing, painting, and printmaking materials and processes | 24 | 57% | 17 | 40% | 1 | 2% | 0 | 0% |
| • Describes drawing, painting, and printmaking processes | 19 | 45% | 22 | 52% | 1 | 2% | 0 | 0% |
| • Solves problems and evaluates possible solutions | 23 | 55% | 17 | 40% | 2 | 5% | 0 | 0% |

[12] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| • Compares materials and techniques, and analyzes the compatibility of materials and techniques | 11 | 26% | 24 | 57% | 7 | 17% | 0 | 0% |
| • Recognizes or identifies processes through reproductions | 10 | 24% | 23 | 55% | 9 | 21% | 0 | 0% |
| • **Knows and understands how to use digital photography and image processes.** | **8** | **19%** | **29** | **69%** | **5** | **12%** | **0** | **0%** |
| • Demonstrates basic camera knowledge (camera parts, vocabulary) | 12 | 29% | 25 | 60% | 5 | 12% | 0 | 0% |
| • Demonstrates knowledge of common editing and imaging software | 5 | 12% | 27 | 64% | 10 | 24% | 0 | 0% |
| • Demonstrates knowledge of uploading, downloading, storing common file types, transferring and printing images | 10 | 24% | 28 | 67% | 4 | 10% | 0 | 0% |
| • Knows and understands how to use the process of creating digital images | 7 | 17% | 23 | 55% | 12 | 29% | 0 | 0% |
| • **Understand materials, tools and processes for videography, filmmaking, and installations** | **4** | **10%** | **11** | **26%** | **23** | **55%** | **4** | **10%** |
| • Identifies/describes materials, tools, and processes for videography, filmmaking and installations | 3 | 7% | 11 | 26% | 24 | 57% | 4 | 10% |
| • **Knows and understands how to use sculptural materials and processes.** | **21** | **50%** | **21** | **50%** | **0** | **0%** | **0** | **0%** |
| • Identifies characteristics of materials | 20 | 48% | 22 | 52% | 0 | 0% | 0 | 0% |
| • Identifies similarities and differences among materials | 14 | 33% | 25 | 60% | 3 | 7% | 0 | 0% |
| • Knows vocabulary related to sculptural materials and processes | 21 | 50% | 21 | 50% | 0 | 0% | 0 | 0% |
| • Describes sculptural processes | 17 | 40% | 24 | 57% | 1 | 2% | 0 | 0% |

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| • Solves problems and evaluates possible solutions | 23 | 55% | 16 | 38% | 3 | 7% | 0 | 0% |
| • Compares materials and techniques, and analyzes the compatibility of materials and techniques | 11 | 26% | 23 | 55% | 8 | 19% | 0 | 0% |
| • Recognizes or identifies processes through reproductions | 10 | 24% | 20 | 48% | 12 | 29% | 0 | 0% |
| • **Knows and understands how to use a variety of fiber art materials and processes[13].** | **5** | **12%** | **26** | **62%** | **8** | **19%** | **2** | **5%** |
| • Identifies characteristics of materials | 3 | 7% | 31 | 74% | 7 | 17% | 1 | 2% |
| • Identifies similarities and differences among materials | 3 | 7% | 26 | 62% | 11 | 26% | 2 | 5% |
| • Knows vocabulary related to fiber materials and processes | 8 | 19% | 20 | 48% | 13 | 31% | 1 | 2% |
| • Describes fiber processes | 3 | 7% | 24 | 57% | 13 | 31% | 2 | 5% |
| • Solves problems and evaluates possible solutions | 9 | 21% | 21 | 50% | 10 | 24% | 2 | 5% |
| • Compares materials and techniques, and analyzes the compatibility of materials and techniques | 4 | 10% | 23 | 55% | 13 | 31% | 2 | 5% |
| • Recognizes or identifies processes through reproductions[13] | 3 | 7% | 23 | 55% | 12 | 29% | 3 | 7% |

[13] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| • **Understands the physical aspects and effective ways of presenting art work for display purposes.** | **20** | **48%** | **18** | **43%** | **4** | **10%** | **0** | **0%** |
| • Identifies and recognizes methods of mounting and matting work in ways appropriate to the medium | 17 | 40% | 22 | 52% | 3 | 7% | 0 | 0% |
| • Identifies and describes methods of displaying three-dimensional work | 19 | 45% | 16 | 38% | 7 | 17% | 0 | 0% |
| • Describes appropriate ways of using exhibition spaces | 15 | 36% | 21 | 50% | 6 | 14% | 0 | 0% |
| **II. A. HISTORICAL & THEORETICAL FOUNDATION OF ART – MATERIALS & PROCESSES** | | | | | | | | |
| • **Understands the following materials within an art historical context: Painting, Drawing, Printmaking, Sculpture, Architecture, Photography, Fiber Arts, Crafts.** | **26** | **62%** | **16** | **38%** | **0** | **0%** | **0** | **0%** |
| • Identifies characteristics of materials, processes, and techniques within an art historical context | 18 | 43% | 24 | 57% | 0 | 0% | 0 | 0% |
| • Identifies similarities and differences among materials, processes, and techniques (e.g., evolution over time) | 15 | 36% | 25 | 60% | 2 | 5% | 0 | 0% |
| • Knows vocabulary related to two-dimensional and three-dimensional media and processes within an art historical context | 24 | 57% | 18 | 43% | 0 | 0% | 0 | 0% |
| • Recognizes or identifies processes within an art historical context through reproductions | 13 | 31% | 27 | 64% | 2 | 5% | 0 | 0% |

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| **II. B. HISTORICAL & THEORETICAL FOUNDATION OF ART – THE WESTERN TRADITION IN ART HISTORY** | | | | | | | | |
| • **Recognizes stylistic traits of art and architecture from each of the following time periods: Prehistory; Ancient Near East; Ancient Greece and Rome; Early Christian, Byzantine and Medieval periods; the Renaissance; the Baroque; 18th through 20th centuries in Europe and North America; contemporary art.** | **24** | **57%** | **17** | **40%** | **1** | **2%** | **0** | **0%** |
| • Identifies the styles of works of art and architecture | 20 | 48% | 22 | 52% | 0 | 0% | 0 | 0% |
| • Categorizes art and architecture according to style and/or period | 17 | 40% | 22 | 52% | 3 | 7% | 0 | 0% |
| • Identifies major works of art and architecture by title, style, and/or artist, as appropriate | 17 | 40% | 21 | 50% | 4 | 10% | 0 | 0% |
| • Analyzes/explains the influence of art periods or schools on later work | 15 | 36% | 26 | 62% | 1 | 2% | 0 | 0% |
| • Analyzes compositional elements and principles of design in works of art and architecture | 21 | 50% | 20 | 48% | 1 | 2% | 0 | 0% |
| • Recognizes the impact of major artistic and technological innovations on the stylistic traits of art | 15 | 36% | 25 | 60% | 2 | 5% | 0 | 0% |

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| • **Understands the content, context, and/or purpose of art and architecture from each of the following time periods: Prehistory; Ancient Near East; Ancient Greece and Rome; Early Christian, Byzantine and Medieval periods; the Renaissance; the Baroque; 18th through 20th centuries in Europe and North America; contemporary art[14].** | **20** | **48%** | **21** | **50%** | **0** | **0%** | **0** | **0%** |
| • Explains the purposes of works of art from various time periods | 15 | 36% | 25 | 60% | 2 | 5% | 0 | 0% |
| • Decodes/analyzes the narrative or intended content of a work of art | 14 | 33% | 25 | 60% | 3 | 7% | 0 | 0% |
| • Analyzes/explains the interrelationships between art and social factors, cultural context, and events | 17 | 40% | 23 | 55% | 2 | 5% | 0 | 0% |
| • Explains the impact of major artistic and technological innovations on the content, context, and purposes of art | 17 | 40% | 20 | 48% | 5 | 12% | 0 | 0% |
| • Acquires and evaluates information about art and artists from various sources | 18 | 43% | 19 | 45% | 5 | 12% | 0 | 0% |

[14] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| **II. C. HISTORICAL & THEORETICAL FOUNDATION OF ART – ART BEYOND THE WESTERN TRADITION** | | | | | | | | |
| • **Knows and understands the general visual characteristics of art and architecture from Asia, Africa, the Americas, the South Pacific region.** | **16** | **38%** | **26** | **62%** | **0** | **0%** | **0** | **0%** |
| • Classifies works of art and architecture by regions/cultures[15] | 13 | 31% | 27 | 64% | 1 | 2% | 0 | 0% |
| • Describes/analyzes works of art and architecture using compositional elements and principles of design | 18 | 43% | 23 | 55% | 1 | 2% | 0 | 0% |
| • Describes/analyzes the interrelationships between art from beyond the Western traditions and art from the Western tradition | 12 | 29% | 25 | 60% | 5 | 12% | 0 | 0% |
| • Identifies major works of art and architecture by title, style, and/or artist, as appropriate | 12 | 29% | 20 | 48% | 10 | 24% | 0 | 0% |

[15] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| • **Understands the general content, context, and purposes of art from Asia, Africa, the Americas, the South Pacific region[16].** | **17** | **40%** | **23** | **55%** | **1** | **2%** | **0** | **0%** |
| • Explains the content and/or purpose (as appropriate) of frequently referenced works of art from various locations and cultures[16] | 15 | 36% | 24 | 57% | 2 | 5% | 0 | 0% |
| • Identifies the general role of a work of art in its culture | 19 | 45% | 21 | 50% | 2 | 5% | 0 | 0% |
| • Explains how the context in which a work of art is created conveys information about various lifestyles and belief systems | 18 | 43% | 21 | 50% | 2 | 5% | 1 | 2% |
| • Acquires and evaluates information about art and artists from various sources | 19 | 45% | 15 | 36% | 8 | 19% | 0 | 0% |
| **II. D. HISTORICAL & THEORETICAL FOUNDATION OF ART – RESPONDING TO ART** | | | | | | | | |
| • **Understands the major theories of art and aesthetics.** | **17** | **40%** | **24** | **57%** | **1** | **2%** | **0** | **0%** |
| • Describes the major characteristics of various theories of art and aesthetics | 14 | 33% | 23 | 55% | 5 | 12% | 0 | 0% |
| • Distinguishes among the major theories of art and aesthetics | 12 | 29% | 25 | 60% | 5 | 12% | 0 | 0% |
| • Compares and contrasts the differences/similarities among theories of art and aesthetics | 10 | 24% | 26 | 62% | 6 | 14% | 0 | 0% |
| • Interprets and evaluates works of art based on theories of art and aesthetics (as opposed to personal opinion) | 17 | 40% | 22 | 52% | 3 | 7% | 0 | 0% |

[16] One or more panelists did not rate this knowledge/skill category or statement.

**Table D6 (continued)**

**Specification Judgments — Combined Across Panels**

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| • **Understands the relationship between art and critical response[17].** | **23** | **55%** | **16** | **38%** | **1** | **2%** | **0** | **0%** |
| • Demonstrates knowledge of critical reactions to well-known works and/or art movements[17] | 14 | 33% | 23 | 55% | 4 | 10% | 0 | 0% |
| • Recognizes/uses multiple viewpoints in examining a work of art | 24 | 57% | 16 | 38% | 2 | 5% | 0 | 0% |
| • Recognizes the way personal experience affects interpretation of art | 25 | 60% | 15 | 36% | 2 | 5% | 0 | 0% |
| • Recognizes and discusses how meaning is created in art (e.g., through symbols, iconography, formal elements and principles) | 28 | 67% | 11 | 26% | 3 | 7% | 0 | 0% |

[17] One or more panelists did not rate this knowledge/skill category or statement.

**Table D7**

**Final Evaluation — Panel 1**

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **Percent** | **N** | **Percent** | **N** | **Percent** | **N** | **Percent** |
| • I understood the purpose of this study. | 17 | 81% | 4 | 19% | 0 | 0% | 0 | 0% |
| • The instructions and explanations provided by the facilitators were clear. | 21 | 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. | 20 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| • The explanation of how the recommended cut score is computed was clear. | 15 | 71% | 6 | 29% | 0 | 0% | 0 | 0% |
| • The opportunity for feedback and discussion between rounds was helpful. | 16 | 76% | 5 | 24% | 0 | 0% | 0 | 0% |
| • The process of making the standard setting judgments was easy to follow. | 13 | 62% | 7 | 33% | 0 | 0% | 1 | 5% |

**Table D7 (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 | Percent | N | Percent | N | Percent |
| • The definition of the JQC | 20 | 95% | 1 | 5% | 0 | 0% |
| • The between-round discussions | 9 | 43% | 11 | 52% | 1 | 5% |
| • The knowledge/skills required to answer each test question | 17 | 81% | 4 | 19% | 0 | 0% |
| • My own professional experience | 10 | 48% | 9 | 43% | 2 | 10% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | Percent | N | Percent | N | Percent | N | Percent |
| • Overall, how comfortable are you with the panel's recommended cut scores? | 18 | 86% | 3 | 14% | 0 | 0% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| | N | Percent | N | Percent | N | Percent |
| • Overall, the recommended cut score is: | 0 | 0% | 21 | 100% | 0 | 0% |

**Table D8**

**Final Evaluation — Panel 2**

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **Percent** | **N** | **Percent** | **N** | **Percent** | **N** | **Percent** |
| • I understood the purpose of this study. | 21 | 100% | 0 | 0% | 0 | 0% | 0 | 0% |
| • The instructions and explanations provided by the facilitators were clear. | 20 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| • The training in the standard setting method was adequate to give me the information I needed to complete my assignment. | 19 | 90% | 2 | 10% | 0 | 0% | 0 | 0% |
| • The explanation of how the recommended cut score is computed was clear. | 15 | 71% | 6 | 29% | 0 | 0% | 0 | 0% |
| • The opportunity for feedback and discussion between rounds was helpful. | 18 | 86% | 3 | 14% | 0 | 0% | 0 | 0% |
| • The process of making the standard setting judgments was easy to follow. | 18 | 86% | 3 | 14% | 0 | 0% | 0 | 0% |

**Table D8 (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 | Percent | N | Percent | N | Percent |
| • The definition of the JQC | 16 | 76% | 5 | 24% | 0 | 0% |
| • The between-round discussions | 13 | 62% | 7 | 33% | 1 | 5% |
| • The knowledge/skills required to answer each test question | 15 | 71% | 6 | 29% | 0 | 0% |
| • My own professional experience | 13 | 62% | 8 | 38% | 0 | 0% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | Percent | N | Percent | N | Percent | N | Percent |
| • Overall, how comfortable are you with the panel's recommended cut scores? | 14 | 67% | 7 | 33% | 0 | 0% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| | N | Percent | N | Percent | N | Percent |
| • Overall, the recommended cut score is: | 1 | 5% | 19 | 90% | 1 | 5% |

# Multistate Standard-Setting Technical Report for the

# *Praxis*® Biology (5236)

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

November 2021

Copyright © 2021 by ETS. All rights reserved. ETS, the ETS logo, and PRAXIS are registered trademarks of ETS. 449059763

# Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Biology (5236) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

## Participating States

Panelists from 14 states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either teachers of biology students or college faculty who prepare those biology teachers and (b) familiarity with the knowledge and skills required of beginning biology 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* Biology test, the recommended passing score[1] is 75 out of a possible 120 raw-score points. The scale score associated with a raw score of 75 is 154 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*® Biology (5236) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in November 2021. Education agencies [2] recommended panelists with (a) experience as either biology teachers and (b) familiarity with the knowledge and skills required of beginning biology teachers. Fourteen states (Table 1) were represented by 21 panelists. (See Appendix A for the names and affiliations of the panelists.)

**Table 1**
***Participating States and the Number of Panelists***

| | |
|---|---|
| Delaware (2 panelists) | Nevada (1 panelist) |
| Idaho (2 panelists) | Rhode Island (1 panelist) |
| Indiana (2 panelists) | Tennessee (2 panelists) |
| Kansas (1 panelist) | Utah (1 panelist) |
| Maryland (2 panelists) | Virginia (2 panelists) |
| North Carolina (2 panelists) | Vermont (1 panelist) |
| New Jersey (1 panelist) | West Virginia (1 panelist) |

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* Biology 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

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

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.

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* Biology 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* Biology 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*® Biology Test

The *Praxis*® Biology *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level biology teachers have the knowledge/skills believed necessary for competent professional practice.

The 2-hour and 30 minute assessment contains 150 selected-response[3] items covering five content areas: *Nature and Impact of Science and Engineering* (approximately 19 items), *Cell Biology: Cell*

[3] Thirty of the 150 selected-response items are pretest items and do not contribute to a candidate's score.

*Structure and Function* (approximately 33 items), *Genetics and Evolution* (approximately 39 items), *Diversity of Life and Organismal Biology* (approximately 20 items), and *Ecology: Organisms and Environments* (approximately 29 items).[4] The reporting scale for the *Praxis* Biology test ranges from 100 to 200 scale-score points.

# Processes and Methods

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

For each panel, the standard-setting study began with a welcome and introduction by the meeting facilitator. The facilitator described the test, provided an overview of standard setting, and presented the agenda for the study. Appendix B shows the 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 biology teachers or areas that address content particularly important for entry-level biology teachers.

## Defining the Just-Qualified Candidate

Following the review of the test, panelists described the just-qualified candidate. The *just-qualified candidate description* plays a central role in standard setting (Perie, 2008); the goal of the standard-setting process is to identify the test score that aligns with this description.

The panelists created a description of the just-qualified candidate, focusing on the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panelists from both panels were assigned to three smaller groups in order to create a

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

draft description. Then they reconvened and, through whole-group discussion of the three drafts, reached consensus on to determine the final version. This final description of the just-qualified candidate was used by both panels for the remainder of the study.

The description of the just-qualified candidate summarized the panels' 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. A clean, PDF-version of the final description was distributed to panelists to use for the remaining phases of the study (see Appendix C for the just-qualified candidate description).

Given that the two-panel multistate standard-setting study was designed to provide two recommendations for the same performance standard, it was important that panels use a consistent just-qualified candidate description to frame their judgments. Therefore, the panelists from both panels worked together until the just-qualified candidate description was finalized.

## Panelists' Judgments

The standard-setting process for the *Praxis* Biology 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. They then 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. If panelists had any questions, those questions were addressed and retraining was provided, if necessary, before proceeding to the first round of judgments.

Following this first round of judgments (*Round 1*), panel-level summaries and item-level feedback were provided to each panel. For each panel, the panelists' judgments were displayed for each item and summarized across the panelists. Item-level data were highlighted to show when panelists converged in their judgments or diverged in their judgments (i.e., when at least two-thirds of the panelists' judgments were Each panel 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 for panelists to conform to another's judgment, but for them to understand the different relevant perspectives among the panelists.

During Round 2, each panel discussed their Round 1 judgments and were encouraged by the facilitator to (a) share the rationales for their judgments and (b) 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 their Round 1 judgment. For each panel, the final judgments for the study, therefore, consist of the panelists' 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 and vice versa. The feedback (data), item-level judgments, and resulting discussions for each panel were independent from the other panel.

# Results

## Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panel included 21 educators representing 14 states. (See Appendix A for a listing of panelists.) Sixteen panelists were teachers, one was an administrator or department head, three were college faculty, and one was a district secondary science academic coach. All of the educators currently, or previously, taught biology. All of the faculty members' job responsibilities included the training of biology 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)***

| Background Survey Question | Number | Percent |
|---|---|---|
| **What is your current position?** | **N** | **%** |
| Teacher | 16 | 76 |
| Administrator or Department head | 1 | 5 |
| College faculty | 3 | 14 |
| District Secondary Science Academic Coach | 1 | 5 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** |
| Black or African American | 4 | 19 |
| Hispanic or Latino | 1 | 5 |
| White | 14 | 67 |
| Prefer not to respond | 2 | 10 |
| **What is your gender?** | **N** | **%** |
| Female | 13 | 62 |
| Male | 8 | 38 |
| **Are you currently certified as a biology teacher in your state?** | **N** | **%** |
| Yes | 20 | 95 |
| No | 1 | 5 |
| **Are you currently teaching biology in your state?** | **N** | **%** |
| Yes | 15 | 71 |
| No | 6 | 29 |
| **Are you currently supervising or mentoring other biology teachers?** | **N** | **%** |
| Yes | 14 | 67 |
| No | 7 | 33 |

*(table continues on the next page)*

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

| **At what P–12 grade level are you currently teaching biology?** | **N** | **%** |
|---|---|---|
| Middle school (6–8 or 7–9) | 3 | 14 |
| High school (9–12 or 10–12) | 10 | 48 |
| Other | 2 | 10 |
| Not currently teaching at the P–12 level | 6 | 29 |
| **Including this year, how many years of experience do you have teaching biology?** | **N** | **%** |
| 3 years or less | 0 | 0 |
| 4–7 years | 5 | 24 |
| 8–11 years | 5 | 24 |
| 12–15 years | 2 | 10 |
| 16 years or more | 9 | 43 |
| **Which best describes the location of your P–12 school?** | **N** | **%** |
| Urban | 3 | 14 |
| Suburban | 8 | 38 |
| Rural | 7 | 33 |
| Not currently working at the P–12 level | 3 | 14 |
| **If you are college faculty, are you currently involved in the training/ preparation of candidates in this subject?** | **N** | **%** |
| Yes | 3 | 14 |
| No | 0 | 0 |
| Not college faculty | 18 | 86 |

## Standard-Setting Judgments

Table 3 summarizes the standard-setting judgments of each panel. Panelist-level results, for Rounds 1 and 2, are presented in Appendix D (Tables D2 – D4). The means represents each panel's passing score recommendation after Round 2. Table 4 also includes each panel's standard deviations (SD) and the standard errors of judgment (SEJ). 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. The confidence intervals created by adding/subtracting two SEJs to each panel's recommended passing score overlap, indicating that they may be comparable. (Appendix E provides the technical notes, which further describe the SEJ.)

**Table 4**

***Summary of Round 2 Standard-setting Judgments by Panel***

| Statistic | Panel 1 | Panel 2 |
|---|---|---|
| Mean | 74.85 | 74.97 |
| Minimum | 61.65 | 62.25 |
| Maximum | 91.10 | 85.65 |
| SD | 8.98 | 7.32 |
| SEJ | 2.84 | 2.21 |

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 Appendix D, Table D4).

The Round 2 mean score is the panel's final recommended passing score. The panel's passing score recommendation for the *Praxis* Biology test are 74.85 for Panel 1 and 74.97 for Panel 2 (out of a possible 120 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended passing score--75 for both Panel 1 and Panel 2. The scale score associated with 75 raw points is 154.

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* Biology test is 74.91 (out of a possible 120 raw-score points). The value was rounded to 75 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 75 raw points is 154.

The conditional standard error of measurement (CSEM) around the recommended passing score is 5.33 raw points. A standard error represents the uncertainty associated with a test score (See Appendix E 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 120 | Praxis Scale Score Equivalent |
|---|---|---|
| RPS - 2 CSEM | 65 | 142 |
| RPS - 1 CSEM | 70 | 148 |
| **RPS** | **75** | **154** |
| RPS +1 CSEM | 81 | 161 |
| RPS +2 CSEM | 86 | 167 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 5.33 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, per panel, is presented in Appendix D (Tables D5 – D12).

All panelists strongly agreed that they understood the purpose of the study. All panelists strongly agreed or agreed that the facilitator's instructions and explanations were clear. All panelists strongly 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 very influential in guiding their standard-setting judgments. All of the panelists reported that round 2 discussions were at least somewhat influential in guiding their judgments. Thirteen of the 21 panelists indicated that their own professional experience was very influential in guiding their judgments.

Thirteen of the 21 panelists indicated they were *very comfortable* with the passing score they recommended. One panelist indicated that they were *very uncomfortable* with the passing score but also indicated that the score was *about right*. All but four of the panelists indicated the recommended

passing score was *about right*; two panelists felt that the passing score was *too low* and two felt that it was *too high*.

# Summary

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Biology 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* Biology test, the recommended passing score[5] is 75 out of a possible 120 raw-score points. The scale score associated with a raw score of 75 is 154 on a 100–200 scale.

[5] 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.

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: Biology (5236)*. 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. (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.

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 and State***

| **Panelist Name** | **Panelists' Affiliation and State Abbreviation** |
|---|---|
| Shannon Bankston-Bell | Charlotte Mecklenburg Schools (NC) |
| Teresa Barton | PikeView High School (WV) |
| Lori Beasley | Seaford School District (DE) |
| Carrie Evick | Cape Henlopen School District (DE) |
| Rebecca Feekes | West Ada School District (ID) |
| Erika Frazier | Prince William County Public Schools (VA) |
| Lori Fretta | Vermont Virtual Learning Cooperative (VT) |
| Christy Gum | Gloucester High School (VA) |
| Will Hemminger | Western Alamance High School (NC) |
| Jenna Henry | Carroll County Public Schools (MD) |
| Jeremy Kelly | Giles County High School (TN) |
| James McCormick | North Plainfield High School (Harrison Bridge) (NJ) |
| Kawonia Mull | Tennessee State University/Williamson County School (TN) |
| Amanda Noble | Portsmouth High School (RI) |
| Benjamin Ott | Boise School District (Les Bois Junior High) (ID) |
| Elvia Solis | Indianapolis Public Schools (IN) |
| Cherry Steffen | Washburn University (KS) |
| Joshua Stowers | Brigham Young University (UT) |
| Brent Strong | Haiser Jr Sr High School (IN) |
| Sheikisha Thomas | Bowie High School (MD) |
| Brian Zeiszler | Great Basin College (NV) |

# Appendix B: Agenda

## *Praxis*® Biology (5236)

## Standard-Setting Study

Day 1 Agenda

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Biology Test

Review the *Praxis* Biology Test

AM Break

Discuss the *Praxis* Biology 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 (small-group drafts) *(continued)*

Collect Materials; End of Day 1

# *Praxis*® Biology (5236)

# Standard-Setting Study

Day 2 Agenda

Overview of Day 2

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

AM Break - Split into two panels

Standard Setting Training in the Modified Angoff Method

Practice Round – Independent Judgments

Lunch

Practice Round –Discussion

Round 1 Standard Setting Judgments

PM Break

Round 1 Standard Setting Judgments *(continued)*

Collect Materials; End of Day 2

# *Praxis*® Biology (5236)

# Standard-Setting Study

Day 3 Agenda

Overview of Day 3

Honoraria Presentation

Round 1 Feedback – Panel Summary

Round 1 Feedback and Round 2 Judgments

AM Break

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

Lunch

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[6]**

A just-qualified candidate…

***Tasks of Teaching Science / Science and Engineering Practices***

1. Is familiar with inquiry-based instructional methods, scaffolding, and sequencing.
2. Understands how to use basic scientific formulas, convert units of measurements, and use proportional reasoning skills
3. Uses scientific language (vocabulary and definitions), models and basic representations to explain scientific concepts
4. Is familiar with how to analyze student ideas in order to identify misconceptions, lack of understanding and gaps of knowledge and can address these topics

***I. Nature and Impact of Science and Engineering***

5. Understands basic lab safety, general science equipment, and experimental design and the processes involved in scientific inquiry

***II. Cell Biology: Cell Structure and Function***

6. Understands how the structure of water and the four major macromolecules affects their functions and properties
7. Understands the enzyme-substrate model and the functions of enzymes and knows the factors that influences the rate of enzymatic reactions
8. Understands active and passive transport and their role in maintaining homeostasis
9. Understands the relationship between common cellular structures and functions and how they interact to perform cellular processes
10. Knows the major biochemical pathways and how they allow the cell to meet its matter and energy needs (i.e., aerobic respiration, anaerobic respiration, photosynthesis, chemosynthesis)

[6] 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…

11. Knows how cells communicate with each other and respond to internal and external stimuli (i.e., cell signaling pathways, plasmodesmata, gap junctions)
12. Knows the different purposes and outcomes of mitosis and meiosis

***III. Genetics and Evolution***

13. Knows the complementary nature of nucleotide base pairs and the general processes and outcomes of replication, transcription, and translation (e.g., mRNA editing, prokaryotes vs. eukaryotes, basic enzymes)
14. Knows basic types and causes of gene (e.g., frame shift) and chromosomal (e.g., translocation) mutations and is familiar with their effects at the cellular (e.g., Sickle Cell) and organism (e.g., Down Syndrome) levels
15. Understands non-Mendelian patterns of inheritance like sex-linked, multiple alleles, co-dominance, incomplete dominance
16. Understands the mechanism of natural selection and is familiar with other mechanisms of evolution (e.g. genetic drift, geographic/reproductive isolation)
17. Knows how fossil records, molecular, structural, and developmental evidence are used to support the theory of evolution
18. Is familiar with basic biotechnology techniques (e.g., DNA extraction, gel electrophoresis, PCR, transformation)

***IV. Diversity of Life and Organismal Biology***

19. Knows the characteristics of living things (distinguishes viruses from cells) and how living things are classified (e.g., Domains, Kingdoms)
20. Knows the basic specialized parts (e.g., adaptations, organ systems, tissues) of plants and animals that function together to allow for survival

***V. Ecology: Organisms and Environments***

21. Knows how organisms interact and respond to their environment, including animal behavior and organismal adaptations

**Description of the Just-Qualified Candidate (continued)**

A just-qualified candidate…

22. Understands how biotic and abiotic components of an ecosystem influence population dynamics and relationships
23. Knows the differences between the flow of energy and the cycling of matter, including biogeochemical cycles and pyramids of biomass and energy
24. Understands how disturbances, natural or human, will affect ecosystems

# Appendix D: Panel-Specific Results

**Table D1**

***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **What is your current position?** | **N** | **%** | **N** | **%** |
| Teacher | 8 | 80 | 8 | 73 |
| Administrator or Department head | 1 | 10 | 0 | 0 |
| College Faculty | 1 | 10 | 2 | 18 |
| District Secondary Science Academic Coach | 0 | 0 | 1 | 9 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** | **N** | **%** |
| Black or African American | 2 | 20 | 2 | 18 |
| Hispanic or Latino | 1 | 10 | 0 | 0 |
| White | 6 | 60 | 8 | 73 |
| Prefer not to respond | 1 | 10 | 1 | 9 |
| **What is your gender?** | **N** | **%** | **N** | **%** |
| Female | 6 | 60 | 7 | 64 |
| Male | 4 | 40 | 4 | 36 |
| **Are you currently certified as a biology teacher in your state?** | **N** | **%** | **N** | **%** |
| Yes | 10 | 100 | 10 | 91 |
| No | 0 | 0 | 1 | 9 |
| **Are you currently teaching biology in your state?** | **N** | **%** | **N** | **%** |
| Yes | 8 | 80 | 7 | 64 |
| No | 2 | 20 | 4 | 36 |
| **Are you currently supervising or mentoring other biology teachers?** | **N** | **%** | **N** | **%** |
| Yes | 6 | 60 | 8 | 73 |
| No | 4 | 40 | 3 | 27 |

*(table continues on the next page)*

**Table D1 (continued from previous page)**
***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **At what P–12 grade level are you currently teaching biology?** | **N** | **%** | **N** | **%** |
| Middle school (6–8 or 7–9) | 1 | 10 | 2 | 18 |
| High school (9–12 or 10–12) | 6 | 60 | 4 | 36 |
| Other | 1 | 10 | 1 | 9 |
| Not currently teaching biology at the P–12 level | 2 | 20 | 4 | 36 |
| **Including this year, how many years of experience do you have teaching biology?** | **N** | **%** | **N** | **%** |
| 3 years or less | 0 | 0 | 0 | 0 |
| 4–7 years | 2 | 20 | 3 | 27 |
| 8–11 years | 1 | 10 | 4 | 36 |
| 12–15 years | 2 | 20 | 0 | 0 |
| 16 years or more | 5 | 50 | 4 | 36 |
| **Which best describes the location of your P–12 school?** | **N** | **%** | **N** | **%** |
| Urban | 1 | 10 | 2 | 18 |
| Suburban | 3 | 30 | 5 | 45 |
| Rural | 5 | 50 | 2 | 18 |
| Not currently working at the P–12 level | 1 | 10 | 2 | 18 |
| **If you are college faculty, are you currently involved in the training/ preparation of biology teachers?** | **N** | **%** | **N** | **%** |
| Yes | 1 | 10 | 2 | 18 |
| No | 0 | 0 | 0 | 0 |
| Not college faculty | 9 | 90 | 9 | 82 |

**Table D2**

***Panel 1 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 62.50 | 65.80 |
| 2 | 84.65 | 80.65 |
| 3 | 69.70 | 70.10 |
| 4 | 60.25 | 61.65 |
| 5 | 73.70 | 74.20 |
| 6 | 65.20 | 67.20 |
| 7 | 83.10 | 80.80 |
| 8 | 73.80 | 74.70 |
| 9 | 92.40 | 91.10 |
| 10 | 83.65 | 82.25 |

**Table D3**

***Panel 2 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 82.10 | 79.05 |
| 2 | 92.95 | 85.65 |
| 3 | 61.45 | 62.25 |
| 4 | 71.40 | 68.10 |
| 5 | 90.35 | 83.90 |
| 6 | 69.65 | 70.85 |
| 7 | 83.80 | 83.30 |
| 8 | 73.20 | 74.10 |
| 9 | 69.75 | 71.45 |
| 10 | 74.80 | 75.10 |
| 11 | 68.50 | 70.95 |

**Table D4**

***Summary of Standard-setting Judgments by Panel and by Round***

| Statistic | Panel 1, Round 1 | Panel 1, Round 2 | Panel 2, Round 1 | Panel 2, Round 2 |
|---|---|---|---|---|
| Mean | 74.90 | 74.85 | 76.18 | 74.97 |
| Minimum | 60.25 | 61.65 | 61.45 | 62.25 |
| Maximum | 92.40 | 91.10 | 92.95 | 85.65 |
| SD | 10.74 | 8.98 | 9.84 | 7.32 |
| SEJ | 3.39 | 2.84 | 2.97 | 2.21 |

***Table D5: Panel 1 Final Evaluation: 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. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The training and practice in the standard setting method was adequate to give me the information I needed to complete my assignment. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 9 | 90 | 1 | 10 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion for round 2 judgments was helpful. | 9 | 90 | 1 | 10 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 7 | 70 | 3 | 30 | 0 | 0 | 0 | 0 |

***Table D6: Panel 1 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 | 9 | 9 | 1 | 10 |
| Whole group JQC consensus | 0 | 0 | 10 | 100 | 0 | 0 |
| Training and practice for making standard-setting judgments | 0 | 0 | 9 | 90 | 1 | 10 |
| Round 1 judgments (independent) | 0 | 0 | 8 | 80 | 2 | 20 |
| Round 2 judgments (with discussion) | 1 | 10 | 8 | 80 | 1 | 10 |

***Table D7: Panel 1 Final Evaluation: 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 | 10 | 100 | 0 | 0 | 0 | 0 |
| The round 2 discussion | 4 | 40 | 6 | 60 | 0 | 0 |
| The knowledge/skills required to answer each test item | 8 | 80 | 2 | 20 | 0 | 0 |
| The passing scores of other panel members | 1 | 10 | 8 | 80 | 1 | 10 |
| My own professional experience | 7 | 70 | 3 | 30 | 0 | 0 |

***Table D8: Panel 1 Final Evaluation: 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? | 5 | 50 | 5 | 50 | 0 | 0 | 0 | 0 |

***Table D9: Panel 1 Final Evaluation: 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: | 1 | 10 | 7 | 70 | 2 | 20 |

***Table D10: Panel 2 Final Evaluation: 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 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 10 | 91 | 1 | 9 | 0 | 0 | 0 | 0 |
| The training and practice in the standard setting method was adequate to give me the information I needed to complete my assignment. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 10 | 91 | 1 | 9 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion for round 2 judgments was helpful. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 10 | 91 | 1 | 9 | 0 | 0 | 0 | 0 |

***Table D11: Panel 2 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 | 1 | 9 | 10 | 91 | 0 | 0 |
| Whole group JQC consensus | 0 | 0 | 11 | 100 | 0 | 0 |
| Training and practice for making standard-setting judgments | 0 | 0 | 11 | 100 | 0 | 0 |
| Round 1 judgments (independent) | 1 | 9 | 7 | 64 | 3 | 27 |
| Round 2 judgments (with discussion) | 1 | 9 | 10 | 91 | 0 | 0 |

***Table D12: Panel 2 Final Evaluation: 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 | 100 | 0 | 0 | 0 | 0 |
| The round 2 discussion | 9 | 82 | 2 | 18 | 0 | 0 |
| The knowledge/skills required to answer each test item | 8 | 73 | 3 | 27 | 0 | 0 |
| The passing scores of other panel members | 3 | 27 | 4 | 36 | 4 | 36 |
| My own professional experience | 6 | 55 | 5 | 45 | 0 | 0 |

*Table D13: Panel 2 Final Evaluation: 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? | 8 | 73 | 2 | 18 | 0 | 0 | 1 | 9 |

*Table D14: Panel 2 Final Evaluation: 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: | 1 | 9 | 10 | 91 | 0 | 0 |

# Appendix E: 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).

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

## Outlier Analysis

An analysis of the data is conducted per panel. Judgments that are above or below 1.5 times the interquartile range for that panel are identified as outliers (High, 2000). ETS makes recommendations on the removal of *specific* outliers based on the observations of the panel facilitator. The panel facilitator reports whether or not the specified panelist was faithfully participating in the standard-setting process. The decision to accept the panel recommendation with or without the outlier data is solely at the discretion of the state.

## 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):

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

# Multistate Standard-Setting Technical Report for the *Praxis*® Chemistry (5246)

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

May 2021

Copyright © 2021 by ETS. All rights reserved. ETS, the ETS logo, and PRAXIS are registered trademarks of ETS. 449059763

# Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Chemistry (5246) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

## Participating States

Panelists from 15 states, and Washington, D.C., were recommended to serve on one of two panels by their respective education agencies. The education agencies recommended panelists with (a) experience as either chemistry teachers or college faculty who prepare chemistry teachers and (b) familiarity with the knowledge and skills required of beginning chemistry 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* Chemistry test, the recommended passing score[1] is 56 out of a possible 100 raw-score points. The scale score associated with a raw score of 56 is 146 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*® Chemistry (5246) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in May 2021. Education agencies[2] recommended panelists with (a) experience as either chemistry teachers or college faculty who prepare chemistry teachers and (b) familiarity with the knowledge and skills required of beginning chemistry teachers. Fifteen 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 , and Washington, D.C., and the Number of Panelists***

| | |
|---|---|
| Arkansas (1 panelist) | Nevada (1 panelist) |
| Delaware (1 panelist) | New Jersey (1 panelist) |
| Indiana (5 panelists) | North Carolina (1 panelist) |
| Kansas (2 panelists) | South Dakota (2 panelists) |
| Kentucky (1 panelist) | Utah (1 panelist) |
| Louisiana (1 panelist) | Virginia (2 panelists) |
| Maryland (2 panelists) | Washington, D.C. (1 panelist) |
| Montana (1 panelist) | West Virginia (2 panelists) |

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

ETS provides a recommended passing score from the multistate standard-setting study to education agencies. In each 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, which represents the combined judgments of a group 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

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

*Praxis* Chemistry passing score (see Geisinger & McCormick, 2010). A state, and D.C. 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, 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* Chemistry test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows states, and D.C. to recognize that any test score on any standardized test—including a *Praxis* Chemistry 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, and D.C. 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, 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 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, and D.C. needs to consider which decision error is more important to minimize.

# Overview of the *Praxis*® Chemistry Test

The *Praxis*® Chemistry *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level chemistry teachers have the knowledge/skills believed necessary for competent professional practice.

The two-hour and 30 minute assessment contains 125 selected-response items[3] covering five content areas: *Nature and Impact of Science and Engineering* (approximately 17 items), *Principles and Models of Matter and Energy* (approximately 31 items), *Chemical Composition, Bonding, and Structure* (approximately 25 items), *Chemical Reactions and Periodicity* (approximately 29 items), and *Solutions and Acid-Base Chemistry* (approximately 23 items).[4] As described in the test description (ETS, in press), more than 40 percent of questions integrate a *Science and Engineering Practice*, and approximately 25 percent of questions assess content applied to a *Task of Teaching of Science*. The reporting scale for the *Praxis* Chemistry test ranges from 100 to 200 scale-score points.

# Processes and Methods

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

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

[3] Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate's score.
[4] The number of items for each content area may vary slightly from form to form of the test.

The test discussion covered the major content areas being addressed by the test. Panelists were asked to remark on any content areas that would be particularly challenging for entry-level teachers or areas that address content particularly important for entry-level teachers.

## Defining the Just-Qualified Candidate

Following the review of the test, panelists described the just-qualified candidate. The *just-qualified candidate description* plays a central role in standard setting (Perie, 2008); the goal of the standard-setting process is to identify the test score that aligns with this description.

Panel 1 created a description of the just-qualified candidate—the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panel 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).

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

## Panelists' Judgments

The standard-setting process for the *Praxis* Chemistry 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 (i.e., when at least two-thirds of the panelists' judgments were in the same difficulty range).

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 from Panel 2 were independent of judgments from Panel 1.

# Results

## Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panel included 25 educators representing 15 states, and D.C. (See Appendix A for a listing of panelists.) Seventeen panelists were teachers, five were college faculty, one was an administrator or department head, one was a director of field experiences, and one was a chemistry teacher/chair of department. Four of the five faculty members' job responsibilities included the training of chemistry 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)***

| Background Survey Question | Number | Percent |
|---|---|---|
| **What is your current position?** | **N** | **%** |
| Teacher | 17 | 68 |
| Administrator/Department head | 1 | 4 |
| College faculty | 5 | 20 |
| Director of Field Experiences (supervises student teachers) | 1 | 4 |
| Chemistry Instructor and Chair of Department | 1 | 4 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** |
| Asian or Asian American | 2 | 8 |
| Black or African American | 2 | 8 |
| Hispanic or Latino | 1 | 4 |
| White | 20 | 80 |
| **What is your gender?** | **N** | **%** |
| Female | 18 | 72 |
| Male | 7 | 25 |

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

| Background Survey Question | Number | Percent |
|---|---|---|
| **Are you currently certified to teach chemistry in your state?** | N | % |
| Yes | 21 | 84 |
| No | 4 | 16 |
| **Are you currently teaching chemistry in your state?** | **N** | **%** |
| Yes | 19 | 76 |
| No | 6 | 24 |
| **Are you currently supervising or mentoring other chemistry teachers?** | **N** | **%** |
| Yes | 9 | 36 |
| No | 16 | 64 |
| **At what P–12 grade level are you currently teaching chemistry?** | **N** | **%** |
| Middle school (6–8 or 7–9) | 2 | 8 |
| Middle and High school | 1 | 4 |
| High school (9–12 or 10–12) | 14 | 56 |
| Not currently teaching at the P–12 level | 8 | 32 |
| **Including this year, how many years of experience do you have teaching chemistry?** | **N** | **%** |
| 3 years or less | 1 | 4 |
| 4–7 years | 6 | 24 |
| 8–11 years | 3 | 12 |
| 12–15 years | 5 | 20 |
| 16 years or more | 10 | 40 |
| **Which best describes the location of your P–12 school?** | **N** | **%** |
| Urban | 5 | 20 |
| Suburban | 7 | 28 |
| Rural | 6 | 24 |
| Not currently working at the P–12 level | 7 | 28 |
| **If you are college faculty, are you currently involved in the training/ preparation of chemistry teacher candidates?** | **N** | **%** |
| Yes | 4 | 16 |
| No | 1 | 4 |
| Not college faculty | 20 | 80 |

## Standard-Setting Judgments

Table 3 summarizes the mean passing score recommendations after Round 2 from each panel. Table 3 also includes the standard deviation and the standard error of judgment (SEJ) (Brennon, 2002, Tannenbaum & Katz, 2013). 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. (For each panel, the panelists' judgments during Rounds 1 and 2, are presented in Appendix D, Tables D2-D4.)

**Table 3**

***Summary of Round 2 Standard-setting Judgments by Panel***

| Statistic | Panel 1 | Panel 2 |
|---|---|---|
| Mean | 51.54 | 59.44 |
| Minimum | 39.20 | 50.00 |
| Maximum | 59.30 | 69.90 |
| SD | 4.82 | 6.90 |
| SEJ | 1.29 | 2.08 |

With multistate standard-setting studies with two panels, the confidence intervals created by adding/subtracting two SEJs to each panel's recommended passing score typically overlap. For this study, the confidence intervals do not overlap, indicating that their recommendations may not be as comparable.

The Round 2 mean score is each panel's final recommended passing score. The panel's passing score recommendation for the *Praxis* Chemistry test are 51.54 for Panel 1 and 59.44 for Panel 2 (out of a possible 100 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended passing score--52 for Panel 1 and 60 for Panel 2. The scale scores associated with 52 and 60 raw points are 140 and 151, 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* Chemistry test is 55.49 (out of a possible 100 raw-score points). The value was rounded to 56 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 56 raw points is 146.

The conditional standard error of measurement (CSEM) around the recommended passing score is 4.99 raw points. A standard error represents the uncertainty associated with a test score (See Appendix E for further information about the CSEM.) Table 4 shows the raw scores and the scale scores associated with one and two CSEM below and above the recommended passing score (See Appendix D, Tables D5 and D6 for this data, per panel).

**Table 4**

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

| Scores | Raw Score Points out of 100 | Praxis Scale Score Equivalent |
|---|---|---|
| **RPS - 2 CSEM** | 47 | 133 |
| **RPS - 1 CSEM** | 52 | 140 |
| **RPS** | **56** | **146** |
| **RPS +1 CSEM** | 61 | 153 |
| **RPS +2 CSEM** | 66 | 160 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 4.99 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, per panel, is presented in Appendix D (Tables D7 – D14).

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 *very influential* or *somewhat influential* in guiding their standard-setting judgments. All of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. More than three-quarters of the panelists (20 of the 25 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

All of the panelists indicated they were at least *somewhat comfortable* with the passing score recommended by their panel; 17 of the 25 panelists were very comfortable. Twenty-three of the 25 panelists indicated their panel's recommended passing score was *about right*. Two panelists from Panel

1 indicated that the passing score was too low; the Panel 1 recommendation was lower than that from Panel 2.

# Summary

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Chemistry 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* Chemistry test, the recommended passing score[5] is 56 out of a possible 100 raw-score points. The scale score associated with a raw score of 56 is 146 on a 100–200 scale.

[5] 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.

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: Chemistry (5246)*. 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** |
|---|---|
| Lori Beasley | Seaford School District (DE) |
| Jennifer Bland | Greenbrier West High School (WV) |
| Spencer Cody | Edmunds Central School District (SD) |
| Amy Connor | Frederick High School (MD) |
| Kyong Damron | Lawrence Central High School (IN) |
| Susanne Dana | Montgomery County Public Schools (VA) |
| Aaron Dehne | Ed W Clark High School Clark County School District (NV) |
| Diane DeVivo | Warren Hills Regional HS (NJ) |
| Lauren Doran | Athens Drive High School (NC) |
| Sharla Dowding | Black Hills State University (SD) |
| Rosie Easterday | Fort Wayne Community Schools (IN) |
| Ivy Fortmeyer | Rocky Mountain College (MT) |
| Maree Funk | Zachary High School (LA) |
| Patricia Hodison | Washington High School (KS) |
| Patricia Kramolisch | Beville Middle School (VA) |
| Sheena Lawson | University of the Cumberlands (KY) |
| James Lipchock | Washington College (MD) |
| Meredith Moore | Washington International School (DC) |
| Adam Robb | Moundridge High School (KS) |
| Rebecca Sansom | Brigham Young University (UT) |
| Michele Shultz | Lebanon Community School Corporation (IN) |
| Michael Tilley | West Virginia University (WV) |
| Grant Wangila | University of Arkansas at Pine Bluff (AR) |
| Lori White | Cascade High School (IN) |

***Participating Panelists With Affiliation and State (continued from previous page)***

| **Panelist Name** | **Panelists' Affiliation and State Abbreviation** |
|---|---|
| Lionel Zhao | Northwest Allen County Schools (IN) |

# Appendix B: Agenda

## *Praxis*® Chemistry (5246)

## Standard-Setting Study

Day 1 Agenda

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Chemistry Test

Review the *Praxis* Chemistry Test

Discuss the *Praxis* Chemistry Test

Lunch

Panel 1: Define the Knowledge/Skills of a Just-Qualified Candidate (small-group drafts)

Panel 1: Define the Knowledge/Skills of a Just-Qualified Candidate (whole-group consensus)

Panel 2: Understand the Knowledge/Skills of the Just-Qualified Candidate (whole-group discussion)

Collect Materials; End of Day 1

# *Praxis*® Chemistry (5246)

# Standard-Setting Study

Day 2 Agenda

Overview of Day 2

Standard Setting Training in the Modified Angoff Method

Round 1 Standard Setting Judgments

Round 1 Feedback and Round 2 Judgments

Lunch

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[6]**

A just-qualified candidate…

Nature and Impact of Science and Engineering

1. Knows the basic principles of experimental designs (e.g., systematic vs. random error).
2. Is familiar with the most common environmental impact of chemical processes
3. Knows how to interpret models to explain scientific phenomenon
4. Knows how to interpret and analyze graphical representations of data

Principles and Models of Matter and Energy

5. Understands the fundamental classifications and properties of matter
6. Is familiar with the fundamental relationship between matter and energy.
7. Knows how to apply gas laws to predict the properties and behavior of ideal gases.
8. Knows the basic principles of thermodynamics (e.g., heat and enthalpy).
9. Knows atomic structure of the quantum mechanical model as it relates to chemical, physical, and nuclear properties of the atom

Chemical Composition, Bonding, and Structure

10. Is familiar with the basic principles of chemical nomenclature
11. Knows how to write formulas of common compounds and polyatomic ions
12. Knows relationships between percent composition and empirical formulas
13. Is familiar with the relationship between a 2-dimensional representation and its 3-dimensional molecular geometry
14. Knows covalent, ionic, and metallic bonding and its relationship to structure and properties.
15. Knows how to identify the predominant intermolecular forces for a substance.
16. Is familiar with how IMFs affect physical properties.

Chemical Reactions and Periodicity

17. Knows the organization of the elements on the periodic table and how it relates to physical and chemical properties of different elements.
18. Understands simple stoichiometric calculations
19. Knows how to predict the products of basic chemical reactions (e.g., single- and double-displacement reactions)
20. Is familiar with basic principles of kinetics, oxidation-reduction, and electrochemistry

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

Solutions and Acid-Base Chemistry

21. Knows basic calculations of solutions such as concentration, dilution, molarity
22. Is familiar with how colligative properties affect the properties of solutions
23. Is familiar with common applications of chemical equilibrium (Le Chatelier's principle) and basic calculations of equilibrium constants (e.g., Law of Mass Action).
24. Knows characteristics of acid and bases (e.g., strong and weak, electrical conductivity, pH scale)
25. Is familiar with simple acid-base titrations

# Appendix D: Panel-Specific Results

**Table D1**

***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **What is your current position?** | **N** | **%** | **N** | **%** |
| Teacher | 9 | 64 | 8 | 73 |
| Administrator/Department head | 0 | 0 | 1 | 9 |
| College faculty | 3 | 21 | 2 | 18 |
| Director of Field Experiences (supervises student teachers) | 1 | 7 | 0 | 0 |
| Chemistry Instructor and Chair of Department | 1 | 7 | 0 | 0 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** | **N** | **%** |
| Asian or Asian American | 1 | 7 | 1 | 9 |
| Black or African American | 2 | 14 | 0 | 0 |
| Hispanic or Latino | 1 | 7 | 0 | 0 |
| White | 10 | 71 | 10 | 91 |
| **What is your gender?** | **N** | **%** | **N** | **%** |
| Female | 8 | 57 | 10 | 91 |
| Male | 6 | 43 | 1 | 9 |
| **Are you currently certified to teach chemistry in your state?** | **N** | **%** | **N** | **%** |
| Yes | 12 | 86 | 9 | 82 |
| No | 2 | 14 | 2 | 18 |
| **Are you currently teaching chemistry in your state?** | **N** | **%** | **N** | **%** |
| Yes | 12 | 86 | 7 | 64 |
| No | 2 | 14 | 4 | 36 |
| **Are you currently supervising or mentoring other chemistry teachers?** | **N** | **%** | **N** | **%** |
| Yes | 6 | 49 | 3 | 27 |
| No | 8 | 57 | 8 | 73 |

**Table D1 (continued from previous page)**
***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **At what P–12 grade level are you currently teaching chemistry?** | **N** | **%** | **N** | **%** |
| Middle school (6–8 or 7–9) | 1 | 7 | 1 | 9 |
| Middle and High school | 0 | 0 | 1 | 9 |
| High school (9–12 or 10–12) | 8 | 57 | 6 | 55 |
| Not currently teaching at the P–12 level | 5 | 36 | 3 | 27 |
| **Including this year, how many years of experience do you have teaching chemistry?** | **N** | **%** | **N** | **%** |
| 3 years or less | 0 | 0 | 1 | 9 |
| 4–7 years | 2 | 14 | 4 | 36 |
| 8–11 years | 3 | 21 | 0 | 0 |
| 12–15 years | 2 | 14 | 3 | 27 |
| 16 years or more | 7 | 50 | 3 | 27 |
| **Which best describes the location of your P–12 school?** | **N** | **%** | **N** | **%** |
| Urban | 3 | 21 | 2 | 18 |
| Suburban | 4 | 29 | 3 | 27 |
| Rural | 2 | 14 | 4 | 36 |
| Not currently working at the P–12 level | 5 | 36 | 2 | 18 |
| **If you are college faculty, are you currently involved in the training/ preparation of chemistry teacher candidates?** | **N** | **%** | **N** | **%** |
| Yes | 3 | 21 | 1 | 9 |
| No | 0 | 0 | 1 | 9 |
| Not college faculty | 11 | 79 | 9 | 82 |

**Table D2**

***Panel 1 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 40.00 | 39.20 |
| 2 | 41.70 | 44.35 |
| 3 | 50.80 | 51.50 |
| 4 | 51.50 | 51.40 |
| 5 | 51.50 | 51.70 |
| 6 | 52.20 | 51.85 |
| 7 | 53.30 | 53.30 |
| 8 | 53.80 | 52.90 |
| 9 | 54.50 | 54.40 |
| 10 | 55.40 | 50.70 |
| 11 | 55.90 | 53.70 |
| 12 | 56.40 | 55.90 |
| 13 | 57.10 | 51.35 |
| 14 | 64.85 | 59.30 |

**Note:** Data from panelists 1, 2, and 14 were detected as outliers (High, 2000; see Appendix E for technical notes). Their scores are not recommended for removal, however, because they were observed to be following the standard-setting process faithfully.

**Table D3**

***Panel 2 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 44.95 | 50.00 |
| 2 | 45.95 | 51.40 |
| 3 | 46.65 | 52.05 |
| 4 | 54.40 | 54.20 |
| 5 | 57.65 | 59.85 |
| 6 | 59.60 | 59.60 |
| 7 | 61.10 | 60.60 |
| 8 | 62.45 | 62.55 |
| 9 | 64.35 | 64.65 |
| 10 | 69.40 | 69.90 |
| 11 | 72.35 | 69.05 |

Table D4 summarizes each panel's judgments, per round. The Round 1 judgments are made without discussion among the panelists. Therefore, the most variability in judgments 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 both panels.

**Table D4**

***Summary of Standard-setting Judgments by Panel and by Round***

| Statistic | Panel 1, Round 1 | Panel 1, Round 2 | Panel 2, Round 1 | Panel 2, Round 2 |
|---|---|---|---|---|
| Mean | 52.78 | 51.54 | 58.08 | 59.44 |
| Minimum | 40.00 | 39.20 | 44.95 | 50.00 |
| Maximum | 64.85 | 59.30 | 72.32 | 69.90 |
| SD | 6.15 | 4.82 | 9.30 | 6.90 |
| SEJ | 1.64 | 1.29 | 2.81 | 2.08 |

Tables D5 and D6 show, for Panel 1 and 2, respectively, the raw scores and the scale scores associated with one and two CSEM below and above each panel's recommended passing scores. 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.

**Table D5**

***Scores 1 and 2 CSEM Around the RPS from Panel 1***

| Panel 1 Scores | Raw Score Points out of 100 | Praxis Scale Score Equivalent |
|---|---|---|
| **RPS - 2 CSEM** | 42 | 126 |
| **RPS - 1 CSEM** | 47 | 133 |
| **Panel 1 RPS** | **52** | **140** |
| **RPS +1 CSEM** | 58 | 148 |
| **RPS +2 CSEM** | 63 | 155 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 5.02 raw points.

**Table D6**

***Scores 1 and 2 CSEM Around the RPS from Panel 2***

| Panel 2 Scores | Raw Score Points out of 100 | Praxis Scale Score Equivalent |
|---|---|---|
| **RPS - 2 CSEM** | 51 | 139 |
| **RPS - 1 CSEM** | 56 | 146 |
| **Panel 2 RPS** | **60** | **151** |
| **RPS +1 CSEM** | 65 | 158 |
| **RPS +2 CSEM** | 70 | 165 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 4.92 raw points.

***Table D7: Panel 1 Final Evaluation: 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. | 13 | 93 | 1 | 7 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 11 | 79 | 3 | 21 | 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 | 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 between rounds was helpful. | 10 | 71 | 4 | 29 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 10 | 71 | 4 | 29 | 0 | 0 | 0 | 0 |

*Table D8: Panel 1 Final Evaluation: 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 | 13 | 93 | 1 | 7 | 0 | 0 |
| The between-round discussions | 6 | 43 | 8 | 57 | 0 | 0 |
| The knowledge/skills required to answer each test item | 13 | 93 | 1 | 7 | 0 | 0 |
| The passing scores of other panel members | 4 | 29 | 10 | 71 | 0 | 0 |
| My own professional experience | 12 | 86 | 2 | 14 | 0 | 0 |

*Table D9: Panel 1 Final Evaluation: 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 % |
|---|---|---|---|---|---|---|---|---|
| Overall, how comfortable are you with the panel's recommended passing score? | 10 | 71 | 4 | 29 | 0 | 0 | 0 | 0 |

*Table D10: Panel 1 Final Evaluation: 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: | 2 | 14 | 12 | 86 | 0 | 0 |

***Table D11: Panel 2 Final Evaluation: 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 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 11 | 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. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion between rounds was helpful. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 11 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |

*Table D12: Panel 2 Final Evaluation: 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 | 100 | 0 | 0 | 0 | 0 |
| The between-round discussions | 8 | 73 | 3 | 27 | 0 | 0 |
| The knowledge/skills required to answer each test item | 10 | 91 | 1 | 9 | 0 | 0 |
| The passing scores of other panel members | 6 | 55 | 4 | 36 | 1 | 9 |
| My own professional experience | 8 | 73 | 3 | 27 | 0 | 0 |

*Table D13: Panel 2 Final Evaluation: 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 % |
|---|---|---|---|---|---|---|---|---|
| Overall, how comfortable are you with the panel's recommended passing score? | 7 | 64 | 4 | 36 | 0 | 0 | 0 | 0 |

*Table D14: Panel 2 Final Evaluation: 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 | 11 | 100 | 0 | 0 |

# Appendix E: 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).

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

## Outlier Analysis

An analysis of the data is conducted per panel. Judgments that are above or below 1.5 times the interquartile range for that panel are identified as outliers (High, 2000). ETS makes recommendations on the removal of *specific* outliers based on the observations of the panel facilitator. The panel facilitator reports whether or not the specified panelist was faithfully participating in the standard-setting process. The decision to accept the panel recommendation with or without the outlier data is solely at the discretion of the state.

## 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):

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

# Multistate Standard-Setting Technical Report for the *Praxis*® Family and Consumer Sciences (5123) Test

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

December 2023

Copyright © 2023 by ETS. All rights reserved. ETS, the ETS logo, and PRAXIS are registered trademarks of ETS. 449059763

# Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Family and Consumer Sciences (5123) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

## Participating States

Panelists from 10 states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as family and consumer sciences teachers or college faculty who prepare family and consumer sciences teachers and (b) familiarity with the knowledge and skills required of beginning family and consumer sciences 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* Family and Consumer Sciences test, the recommended passing score is 66 out of a possible 110 raw-score points. The scale score associated with a raw score of 66 is 151 on a 100–200 scale.

# Introduction

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Family and Consumer Sciences (5123) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in December 2023. Education agencies[1] recommended panelists with (a) experience as family and consumer sciences teachers or college faculty who prepare those family and consumer sciences teachers and (b) familiarity with the knowledge and skills required of beginning family and consumer sciences teachers. Ten states (Table 1) were represented by 13 panelists. (See Appendix A for the names and affiliations of the panelists.)

**Table 1**
***Participating States and the Number of Panelists***

| | |
|---|---|
| Alabama (2 panelists) | North Carolina (1 panelist) |
| Connecticut (1 panelist) | Pennsylvania (1 panelist) |
| Delaware (1 panelist) | Utah (1 panelist) |
| Kansas (1 panelist) | Wisconsin (1 panelist) |
| Mississippi (2 panelists) | Wyoming (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, 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* Family and Consumer Sciences 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

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

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.

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* Family and Consumer Sciences 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* Family and Consumer Sciences 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*® Family and Consumer Sciences Test

The *Praxis*® Family and Consumer Sciences *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level family and consumer sciences teachers have the knowledge/skills believed necessary for competent professional practice.

The 2-hour and 10 minute assessment contains 130 selected-response items[2] covering six content areas: *Family, Career, and Community Interrelationships* (approximately 22 items), *Wellness, Nutrition, and Food Science* (approximately 27 items), *Human Development and Interpersonal Relationships* (approximately 23 items) ), *Housing, Interior Design, Textiles, and Apparel* (approximately 16 items) ), *Business Skills Across Career Pathways* (approximately 16 items), and *Professional Teaching Practices* (approximately 26 items).[3] The reporting scale for the *Praxis* Family and Consumer Sciences 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 family and consumer sciences teachers or areas that address content particularly important for entry-level family and consumer sciences teachers.

[2] Twenty of the 130 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.

## 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 panelists created a description of the just-qualified candidate, focusing on the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panel first split into three breakout groups to consider the just-qualified candidate. in order to create a draft description. Then they reconvened and, through whole-group discussion of the three drafts, reached consensus on the final version 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. A clean, PDF-version of the final description was distributed to panelists to use for the remaining phases of the study (see Appendix C for the just-qualified candidate description).

## Panelists' Judgments

The standard-setting process for the *Praxis* Family and Consumer Sciences 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 diverged in their judgments or converged in their judgments (i.e., when at least two-thirds of the panelists' judgments were in the same difficulty range).

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 each other's judgment, but for them to understand the different, but relevant, perspectives among them.

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 13 educators representing 10 states. (See Appendix A for a listing of panelists.) Eight panelists were teachers, four were college faculty, and one was a state consultant. Six of the teachers and the state consultant also supervise or mentor other family and consumer sciences teachers. Additionally, all four of the faculty members' job responsibilities included the training of family and consumer sciences teachers.

**Table 2**
***Panel Member Demographics***

| Background Survey Question | Number | Percent |
|---|---|---|
| **What is your current position?** | **N** | **%** |
| Teacher | 8 | 62 |
| College faculty | 4 | 31 |
| Family and Consumer Sciences Education Consultant - State | 1 | 8 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** |
| Black or African American | 2 | 15 |
| White | 10 | 77 |
| Prefer to not respond | 1 | 8 |
| **What is your gender?** | **N** | **%** |
| Female | 13 | 100 |
| **Are you currently certified to teach family and consumer sciences in your state?** | **N** | **%** |
| Yes | 9 | 69 |
| No | 0 | 0 |
| I am not currently working at the P-12 level | 4 | 31 |
| **Are you currently teaching family and consumer sciences in your state?** | N | % |
| Yes | 8 | 62 |
| No | 1 | 8 |
| I am not currently working at the P-12 level | 4 | 31 |
| **Are you currently supervising or mentoring other family and consumer sciences teachers?** | **N** | **%** |
| Yes | 7 | 54 |
| No | 2 | 15 |
| I am not currently working at the P-12 level | 4 | 31 |

*Table continues on the next page.*

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

| **Including this year, how many years of experience do you have teaching family and consumer sciences?** | **N** | **%** |
|---|---|---|
| 3 years or less | 0 | 0 |
| 4–7 years | 1 | 8 |
| 8–11 years | 1 | 8 |
| 12–15 years | 1 | 8 |
| 16 years or more | 6 | 46 |
| I am not currently working at the P-12 level | 4 | 31 |
| **Which best describes the location of your K–12 school?** | **N** | **%** |
| Urban | 2 | 15 |
| Suburban | 3 | 23 |
| Rural | 3 | 23 |
| I am not currently working at the P-12 level | 5 | 38 |
| **If you are college faculty, are you currently involved in the training/ preparation of family and consumer sciences teachers?** | **N** | **%** |
| Yes | 4 | 31 |
| No | 0 | 0 |
| Not college faculty | 9 | 69 |

## Standard-Setting Judgments

Table 3 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 110. Data from Panelists 7 and 10 were detected to be outliers (High, 2000; see Appendix D). However, ETS does not recommend that their data be removed from the panel recommendation. Based on a report from the panel facilitator, the panelists were believed to be following the standard-setting process faithfully. Throughout the standard-setting, panelists are encouraged to consider the perspectives of their colleagues but that were not required to agree with their judgments.

**Table 3**

***Raw Score Recommendation of Each Panelist by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 66.00 | 66.60 |
| 2 | 65.70 | 67.50 |
| 3 | 63.65 | 62.85 |
| 4 | 52.40 | 64.90 |
| 5 | 74.70 | 72.80 |
| 6 | 72.40 | 71.80 |
| 7 | 85.50 | 79.85 |
| 8 | 66.95 | 67.95 |
| 9 | 70.80 | 70.25 |
| 10 | 47.10 | 47.60 |
| 11 | 61.50 | 64.10 |
| 12 | 51.60 | 55.70 |
| 13 | 67.30 | 65.50 |

Table 4 summarizes the standard-setting judgments of the panel at each round of judgment. The mean represents the panel's passing score recommendation at each round. Table 4 also includes the standard deviation 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. 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. Appendix F provides the results without the outliers included.)

**Table 4**

***Summary Statistics by Round of Judgments***

| Statistic | Round 1 | Round 2 |
|---|---|---|
| Mean | 65.05 | 65.95 |
| Minimum | 47.10 | 47.60 |
| Maximum | 85.50 | 79.85 |
| SD | 10.35 | 7.94 |
| SEJ | 2.87 | 2.20 |

Round 1 judgments are made without discussion among the panelists. Therefore, there is typically more variability in judgments 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* Family and Consumer Sciences test is 65.95 (out of a possible 110 raw-score points). The value was rounded to the next highest whole number, 66, to determine the functional recommended passing score. The scale score associated with 66 raw points is 151.

The conditional standard error of measurement (CSEM) around the recommended passing score is 5.16 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 110 | Praxis Scale Score Equivalent |
|---|---|---|
| RPS - 2 CSEM | 56 | 138 |
| RPS - 1 CSEM | 61 | 144 |
| **RPS** | **66** | **151** |
| RPS +1 CSEM | 72 | 158 |
| RPS +2 CSEM | 77 | 165 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 5.16 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 and that the facilitator's instructions and explanations were clear. All panelists *strongly 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. All of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Nine of the 12 panelists indicated that their own professional experience was *very influential* in guiding their judgments.

Ten of the 13 panelists indicated they were *very comfortable* with the passing score they recommended; two panelists indicated that they were *somewhat comfortable,* and one panelist was *somewhat uncomfortable* with the panels recommended passing score. Eleven of the 13 panelists indicated the recommended passing score was *about right*; two panelists indicated that the recommended passing score was *too low*.

# Summary

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Family and Consumer Sciences 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* Family and Consumer Sciences test, the recommended passing score is 66 out of a possible 110 raw-score points. The scale score associated with a raw score of 66 is 151 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: Family and Consumer Sciences (5123)*. Princeton, NJ: Author.

Geisinger, K. F., & McCormick, C. M. (2010), Adopting cut scores: post-standard-setting panel considerations for decision makers. *Educational Measurement: Issues and Practice*, *29*, 38–44.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport, CT: American Council on Education/Praeger.

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

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** |
|---|---|
| Shandi Andres | Kansas State University (KS) |
| Michelle Bassett-Shermo | Eau claire Memorial High School (WI) |
| Torey Berdan | Natrona County School District - Pathways Innovation Center (WY) |
| LaTonya Dixon | Alabama A & M University (AL) |
| Rachel Keleher | Fairfield Warde High School (CT) |
| Jamie Kennedy | Sweetwater County School District #2 (WY) |
| Virginia Kochan-Lovett | Limestone County Schools Alabama (AL) |
| Amy Lloyd | TH Bell Junior High (UT) |
| Linda Marvel | Cape Henlopen School District (DE) |
| Kalani Palmer | Indiana University of PA (PA) |
| Angel Pilcher | Center Hill High School: DeSoto County Schools (MS) |
| Mariann Taylor | The University of Southern Mississippi (MS) |
| Gwen Turner | NC Department of Public Instruction (NC) |

***Note.*** Panelists provided permission to be listed in this report.

# Appendix B: Agenda

## *Praxis*® Family and Consumer Sciences (5123)

## Standard-Setting Study

**DAY 1 AGENDA**

| | |
|---|---|
| **10:00 AM ET/9:00 AM CT/ 8:00 AM MT/ 7:00 AM PT/6:00 AM AK/5:00 AM HI** | Welcome, introductions, and goals for the day<br><br>Standard-setting overview presentation<br>• Q&A about the training<br>• Polling: Checking for understanding<br><br>Test familiarization for the test<br>• Overview & instructions<br>• Independent test familiarization<br>• Self-scoring instructions<br>• Independent self-scoring<br><br>**Break**<br><br>Discussion of the content measured<br><br>**Lunch break**<br><br>Just-Qualified Candidate<br>• Overview<br>• Polling: Evaluation of JQC training<br>• Small group drafts<br>**Break**<br><br>Just-Qualified Candidate (continued)<br>• Whole group consensus |
| **7:00 PM ET/ 6:00 PM CT/ 5:00 PM MT/ 4:00 PM PT/3:00 PM AK/2:00 PM HI** | End of Day 1 |

# *Praxis*® Family and Consumer Sciences (5123)

# Standard-Setting Study

**DAY 2 AGENDA**

| | |
|---|---|
| **10:00 AM ET/9:00 AM CT/ 8:00 AM MT/ 7:00 AM PT/6:00 AM AK/5:00 AM HI** | Overview of the Day<br><br>Honoraria Payment Process Presentation<br><br>Standard Setting Training for Selected-Response Items<br>• Instructions and materials<br>• Independent practice round judgments<br><br>**Break**<br><br>Practice Round Data Discussion<br>• Instructions, materials, and screen setup<br>• Discussion of the practice round data<br>• Polling: Evaluation of the judgment training<br><br>Round 1 Standard Setting Judgments<br><br>**Lunch break**<br><br>Round 1 feedback: Summary data<br>• Polling: Evaluation of data presentation<br>Round 1 feedback: Item-level data and<br>Round 2 judgments (break as needed)<br>• Check out when finished<br>**BREAK**<br><br>Round 2 feedback: Recommended Passing Scores<br><br>Complete final evaluation<br><br>Wrap Up/ Final Steps<br>• Review Nondisclosure of Secure Materials<br>  ○ Destruction of Files<br>  ○ What you can/cannot discuss<br>• Final Questions and Thank You |
| **7:00 PM ET/ 6:00 PM CT/ 5:00 PM MT/ 4:00 PM PT/3:00 PM AK/2:00 PM HI** | End of Day 2 |

# Appendix C: Just-Qualified Candidate Description

**Description of the Just-Qualified Candidate[4]**

A just-qualified candidate…

1. Understands strategies and resources for promoting healthy relationships and communication skills for diverse families and individuals
2. Knows consumer rights, roles and responsibilities related to individual, family and government resources
3. Understands decision-making and problem-solving processes for individual, family, career and consumer issues
4. Understands basic safety and sanitation procedures in all areas of Family and Consumer Sciences
5. Knows general concepts of food science
6. Understands food preparation techniques, knowledge, skills & lab management
7. Knows roles and skills utilized in the hospitality industry
8. Understands basic nutrients and dietary guidelines recommended across the lifespan
9. Knows the elements and principles of design
10. Knows basic sewing tools, construction techniques, care, and design of textile products and apparel
11. Familiar with basic history of housing architecture & related design concepts
12. Understands the foundational theories related to the stages, characteristics, and interrelatedness of physical, social, emotional, moral, and cognitive development across the lifespan
13. Knows the components of a curriculum that promote and address the diverse needs of all students
14. Knows methods for implementing FCS academic standards into authentic, student-centered learning experiences and assessments
15. Is familiar with career paths and professional organizations within FCS disciplines including the Body of Knowledge.
16. Is familiar with culturally competent principles and practices for managing diverse individual and family resources
17. Understands essential financial literacy concepts and their impacts on financial management across the lifespan
18. Understands FCCLA, resources, and integration into the FCS courses
19. Is familiar with how CTE programs and CTSOs are supported through legislation
20. Understands parenting styles, family interactions and their impact on growth and development.

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

# 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).

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

## Outlier Analysis

An analysis of the data is conducted per panel. Judgments that are above or below 1.5 times the interquartile range for that panel are identified as outliers (High, 2000). ETS makes recommendations on the removal of *specific* outliers based on the observations of the panel facilitator. The panel facilitator reports whether or not the specified panelist was faithfully participating in the standard-setting process. The decision to accept the panel recommendation with or without the outlier data is solely at the discretion of the state.

## 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**

***Final Evaluation 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 | 85 | 2 | 15 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 10 | 77 | 3 | 23 | 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 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 9 | 69 | 4 | 31 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion for round 2 judgments was helpful. | 11 | 85 | 2 | 15 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 11 | 85 | 2 | 15 | 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 | 1 | 8 | 12 | 92 | 0 | 0 |
| Whole group JQC consensus | 1 | 8 | 12 | 92 | 0 | 0 |
| Training and practice for making standard-setting judgments | 0 | 0 | 13 | 100 | 0 | 0 |
| Round 1 judgments (independent) | 3 | 23 | 10 | 77 | 0 | 0 |
| Round 2 judgments (with discussion) | 1 | 8 | 10 | 77 | 2 | 15 |

**Table E3**

***Final Evaluation: 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 | 12 | 92 | 1 | 8 | 0 | 0 |
| The between-round discussions | 6 | 46 | 7 | 54 | 0 | 0 |
| The knowledge/skills required to answer each test item | 9 | 69 | 4 | 31 | 0 | 0 |
| The passing scores of other panel members | 0 | 0 | 12 | 92 | 1 | 8 |
| My own professional experience | 9 | 69 | 4 | 31 | 0 | 0 |

**Table E4**

***Final Evaluation: 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 % |
|---|---|---|---|---|---|---|---|---|
| Overall, how comfortable are you with the panel's recommended passing score? | 10 | 77 | 2 | 15 | 1 | 8 | 0 | 0 |

**Table E5**

***Final Evaluation: 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: | 2 | 15 | 11 | 85 | 0 | 0 |

# Appendix F: Results Without Outliers

**Table F1**

***Summary Statistics by Round of Judgments***

| Statistic | Round 1 | Round 2 |
|---|---|---|
| Mean | 64.82 | 66.36 |
| Minimum | 51.60 | 55.70 |
| Maximum | 74.70 | 72.80 |
| SD | 7.38 | 4.73 |
| SEJ | 2.23 | 1.43 |

**Table F2**

***Scores 1 and 2 CSEM Around Study Value Without Outliers (SV)***

| Scores | Raw Score Points out of 110 | Praxis Scale Score Equivalent |
|---|---|---|
| SV - 2 CSEM | 57 | 139 |
| SV - 1 CSEM | 62 | 146 |
| **SV** | **67** | **152** |
| SV +1 CSEM | 73 | 160 |
| SV +2 CSEM | 78 | 166 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the SV is 5.14 raw points. The unrounded CSEM value is added to, or subtracted from, the rounded SV. The resulting values are rounded up to the next-highest whole number and then converted to scale scores.

Listening. Learning. Leading.

# Multi-State Standard Setting Technical Report

# PRAXIS™ MUSIC: CONTENT KNOWLEDGE (0113)

Educational and Credentialing Research

Educational Testing Service

Princeton, New Jersey

November 2011

Copyright © 2011 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.

## Executive Summary

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for the Praxis™ Music: Content Knowledge (0113) test, research staff from Educational Testing Service (ETS) designed and conducted a two-panel, multi-state standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level music teachers.

### Participating States

Panelists from seventeen states and the District of Columbia were recommended by state departments of education to participate on expert panels. The state departments of education recommended panelists with (a) education experience, either as music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music teachers.

### Recommended Cut Scores

The recommended passing score for each panel, as well as the average passing score across the two panels, are provided to help state departments of education determine an appropriate operational passing score. For the Praxis Music: Content Knowledge test, the recommended passing score[1] is 70 (on the raw score metric), which represents 64% of the total available 110 raw score points. The scaled score associated with a raw score of 70 is 161 (on a 100 - 200 scale).

### Summary of Content Specification Judgments

Panelists judged the extent to which the knowledge and skills reflected by the content specifications were important for entry-level music teachers. The favorable judgments of the panelists provided evidence that the content covered by the test is important for beginning practice.

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

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for the Praxis™ Music: Content Knowledge (0113) test, research staff from Educational Testing Service (ETS) designed and conducted a two-panel, multi-state standard-setting study[2]. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level music teachers. Panelists were recommended by state departments of education[3] to participate on the expert panels. The state departments of education recommended panelists with (a) education experience, either as music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music teachers.

The two, non-overlapping panels (a) allow each participating state to be represented and (b) provide a replication of the judgment process to strengthen the technical quality of the recommended passing score. Seventeen states and the District of Columbia (see Table 1) were represented by 39 panelists across the panels. (See Appendix A for the names and affiliations of the panelists.)

**Table 1**

***Participating States and Number of Panelists(Across Panels)***

| | |
|---|---|
| Arkansas (3 panelists) | North Carolina (3 panelist) |
| Connecticut (1 panelist) | New Hampshire (3 panelists) |
| District of Columbia (3 panelists) | Nevada (2 panelists) |
| Hawaii (2 panelists) | Ohio (3 panelists) |
| Idaho (2 panelists) | Pennsylvania (1 panelist) |
| Kentucky (3 panelists) | Rhode Island (1 panelist) |
| Maine (1 panelist) | South Carolina (2 panelists) |
| Maryland (3 panelists) | Tennessee (2 panelists) |
| Missouri (2 panelist) | Vermont (2 panelists) |

[2] The multi-state standard-setting study collected judgments for two related Praxis tests — Praxis Music: Content Knowledge (0113) and Praxis Music: Content and Instruction (0114). Separate technical reports were prepared for each test.

[3] State departments of education that currently use one or more Praxis tests were invited to participate in the multi-state standard-setting study.

The panels were convened in October/November 2011 in Princeton, New Jersey. For both panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing scores.

The following technical report is divided into 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.

The passing-score recommendation for the Praxis Music: Content Knowledge test is provided to each of the represented state departments of education. In each state, the department of education, the state board of education, or a designated educator licensure board is responsible for establishing the final passing score in accordance with applicable state regulations. The study provides a recommended passing score, which represent the combined judgments of two groups of experienced educators. The full range of a state department of education's needs and expectations cannot likely be represented during the standard-setting study. Each state, therefore, may want to consider the recommended passing score (as well as the separate panels' recommended passing scores) and other sources of information when setting the final Praxis Music: Content Knowledge 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.

Two sources of information to consider when setting the passing score are the standard errors of measurement (SEM) and the standard errors of judgment (SEJ). The former addresses the reliability of Praxis Music: Content Knowledge test score and the latter, the reliability of panelists' passing-score recommendations. The SEM allows a state to recognize that a Praxis Music: Content Knowledge test score—any test score on any test—is less than 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 state to consider the likelihood that the recommended passing score from the current panels would be similar to the passing score 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 for a test 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 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 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 needs to consider which decision error may be more important to minimize.

## Overview of the Praxis Music: Content Knowledge Test

The Praxis Music: Content Knowledge *Test at a Glance* document (ETS, 2010) describes the purpose and structure of the test. In brief, the test measures whether entry-level music teachers have the knowledge and skills believed necessary for competent professional practice. A National Advisory Committee of expert practitioners and preparation faculty defined the content of the test, and a national survey of the field confirmed the content.

The two hour assessment contains 120 multiple-choice questions[4] covering four content areas: *Music History and Literature* (approximately 17 questions); *Theory and Composition* (approximately 19 questions); *Performance* (approximately 28 questions); and *Pedagogy, Professional Issues, and Technology* (approximately 56 questions)[5]. The reporting scale for the Praxis Music: Content Knowledge test ranges from 100 to 200 scaled-score points.

[4] Ten of the 120 multiple-choice questions are pretest questions and do not contribute to a candidate's score.
[5] The number of questions for each content area may vary slightly from form to form of the test.

## Processes and Methods

For both expert panels, the same processes and methods were used to train panelists, gather panelists' judgments and to calculate the recommended passing scores. The following section describes the standard-setting processes and methods. (The agenda for the panel meetings are presented in Appendix B.)

The design of the standard-setting study included two non-overlapping expert panels. The training provided to panelists as well as the study materials were consistent across panels with the exception of defining the Just Qualified Candidate (JQC). To assure that both panels were using the same frame of reference when making question-level standard-setting judgments, the JQC definition developed through a consensus process by the first panel was used as the definition for the second panel. The second panel did complete a thorough review of the definition to allow panelists to internalize the definition. The processes for developing the definition (with Panel 1) and reviewing/internalizing the definition (with Panel 2) are described later, and the JQC definition is presented in Appendix C.

The panelists were sent an e-mail explaining the purpose of the standard-setting study and requesting that they review the content specifications for the test (included in the *Test at a Glance* document, which was attached to the e-mail). The purpose of the review was to 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 explained how the test was developed, provided an overview of standard setting, and presented the agenda for the study.

### Reviewing the Test

The first activity was for the panelists to "take the test." (Each panelist had signed a nondisclosure form.) The panelists were given approximately an hour and a half to respond to the multiple-choice questions. (Panelists were instructed not to refer to the answer key while taking the test.) The purpose of "taking the test" was for the panelists to become familiar with the test format, content, and difficulty. After "taking the test," the panelists checked their responses against the answer key.

The panelists then engaged in a discussion of the major content areas being addressed by the test; they were also asked to remark on any content areas that they thought would be particularly challenging

for entering music teachers, and areas that addressed content that would be particularly important for entering music teachers.

**Defining the Just Qualified Candidate**

Following the review of the test, panelists internalized the definition of the Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of knowledge believed necessary to be a qualified music teacher. The JQC definition is the operational definition of the passing score. The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC.

Panel 1 developed the JQC definition. The panelists were split into smaller groups, and each group was asked to write down their definition of a JQC. Each group referred to the Praxis Music: Content Knowledge *Test at a Glance* to guide their definition. Each group posted its definition on chart paper, and a full-panel discussion occurred to reach a consensus on a definition (see Appendix C for the definition).

For Panel 2, the panelists began with the definition of the JQC developed by Panel 1. Given that the multi-state standard-setting study was designed to replicate processes and procedures across the two panels, it was important that both panels use consistent JQC definitions to frame their judgments. The panelists reviewed the JQC definition, and any ambiguities were discussed and clarified. The panelists then were split into smaller groups, and each group developed performance indicators or "can do" statements based on the definition. The purpose of the indicators was to provide clear examples of what might be observed to indicate that the teacher had the defined knowledge and skills. The performance indicators were shared and discussed.

**Panelists' Judgments**

The standard-setting process for the Praxis Music: Content Knowledge test was a probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC would answer it 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 a JQC would answer the question

correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly.

For both panels, the panelists were asked to approach the judgment process in two stages. First, they reviewed the definition of the JQC and the question and decided if, overall, the question was difficult for the JQC, easy for the JQC, or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rule of thumb to guide their decision:

- difficult questions for a JQC were in the 0 to .30 range;
- moderately difficult/easy questions for a JQC were in the .40 to .60 range; and
- easy questions for a JQC were in the .70 to 1 range.

The second decision was for panelists to decide how they wanted to refine their judgment within the range. For example, if a panelist thought that a question was easy for a JQC, the initial decision located the question in the .70 to 1 range. The second decision was for the panelist to decide if the likelihood of answering it correctly was .70, .80, .90, .95, or 1.0. The two-stage decision-process was implemented to reduce the cognitive load placed on the panelists. The panelists practiced making their standard-setting judgments on four questions on the test.

The panelists engaged in two rounds of judgments. Following Round 1, question-level feedback was provided to the panel. The panelists' judgments were displayed for each question. The panelists' judgments were summarized by the three general difficulty levels (0 to .30, .40 to .60, and .70 to 1), and the panel's average question judgment was provided. Questions were highlighted to show when panelists converged in their judgments (at least two-thirds of the panelists located a question in the same difficulty range) or diverged in their judgments. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2).

Other than the definition of the JQC, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

### Judgment of Content Specifications

In addition to the two-round standard-setting process, each panel judged the importance of the knowledge and skills stated or implied in the content specifications for the job of an entry-level music teacher. These judgments addressed the perceived content-based validity of the test. Judgments were made using a four-point scale — V*ery Important*, I*mportant*, S*lightly Important*, and *Not Important*. Each panelist independently judged the knowledge categories and knowledge statements.

## Results

The recommended passing score presented is the average of the results from the two panels. Results from the separate panels also are presented. More detailed results are presented in Appendix D.

### Expert Panels

The two panels that comprised the study included 39 educators representing seventeen states and the District of Columbia. (See Appendix A for a listing of panelists.) In brief, 28 panelists were teachers and 11 were college faculty. All of the panelists who were college faculty were currently involved in the training or preparation of music teachers. Thirty-one panelists were White, five were Black or African American, one was Hispanic or Latino, one was Asian or Asian American, and one was Multiracial. Twenty-one panelists were female. More than half of the panelists (22 of the 39 panelists) had 11 or fewer years of experience as a teacher.

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

**Table 2**

***Panel Member Demographics (Across Panels)***

| | *N* | % |
|---|---|---|
| **Current Position** | | |
| Teacher | 28 | 72% |
| College Faculty | 11 | 28% |
| **Race** | | |
| White | 31 | 79% |
| Black or African American | 5 | 13% |
| Hispanic or Latino | 1 | 3% |
| Asian or Asian American | 1 | 3% |
| Multiracial | 1 | 3% |
| **Gender** | | |
| Female | 21 | 54% |
| Male | 18 | 46% |
| **Which of the following best describes your music education specialty?** | | |
| General Music Education | 14 | 36% |
| Instructional Music Education | 17 | 44% |
| Vocal Music Education | 6 | 15% |
| Other | 2 | 5% |
| **Are you currently certified as a music teacher in your state?** | | |
| Yes | 32 | 82% |
| No | 7 | 18% |
| **At what K-12 grade level are you currently a music teacher?** | | |
| Elementary | 8 | 21% |
| Middle School | 7 | 18% |
| Elementary and Middle School | 3 | 8% |
| High School | 5 | 13% |
| Middle and High School | 3 | 8% |
| All Grades | 1 | 3% |
| Not currently a music teacher at the K-12 level | 12 | 31% |

**Table 2 (continued)**

***Panel Member Demographics (Across Panels)***

| | *N* | % |
|---|---|---|
| **How many years of experience do you have teaching?** | | |
| 3 years or less | 5 | 13% |
| 4 - 7 years | 6 | 15% |
| 8 - 11 years | 11 | 28% |
| 12 - 15 years | 6 | 15% |
| 16 years or more | 11 | 28% |
| **Which best describes the location of your K-12 school?** | | |
| Urban | 11 | 28% |
| Suburban | 8 | 21% |
| Rural | 9 | 23% |
| Not currently working at the K-12 level | 11 | 28% |
| **If you are college faculty, are you currently involved in the training/preparation of music teachers?** | | |
| Yes | 11 | 28% |
| No | 0 | 0% |
| Not college faculty | 28 | 72% |

**Initial Evaluation Forms**

The panelists completed an initial evaluation after receiving training on how to make standard-setting judgments. The primary information collected from this form was the panelists indicating if they had received adequate training to make their standard-setting judgments and were ready to proceed. Across both panels, all panelists indicated that they were prepared to make their judgments.

**Summary of Standard-setting Judgments**

A summary of standard-setting judgments (Round 2) are presented in Table 3. The numbers in the table summarize the recommended passing scores—the number of raw points needed to "pass" the test. The panel's average recommended passing score and highest and lowest passing scores are reported, as are the standard deviations (SD) of panelists' passing scores and the standard errors of judgment (SEJ). Panelist-level results, for Rounds 1 and 2, are presented in Appendix D (see Tables D2 and D3).

The SEJ is one way of estimating the reliability of the judgments[6]. It indicates how likely it would be for 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. A comparable panel's passing score would be within one SEJ of the current average passing score 68 percent of the time.

The panels' passing score recommendations for the Praxis Music: Content Knowledge test are 71.96 for Panel 1 and 67.78 for Panel 2 (out of a possible 110 raw-score points). The values were rounded to the next highest whole number to determine the functional recommended cut scores — 72 for Panel 1 and 68 for Panel 2. The scaled scores associated with 72 and 68 raw points are 164 and 159, respectively.

**Table 3**

***Summary of Round 2 Standard-setting Judgments***

| | Panel 1 | Panel 2 |
|---|---|---|
| **Average** | 71.96 | 67.78 |
| **Median** | 72.00 | 68.90 |
| **Lowest** | 49.35 | 56.95 |
| **Highest** | 84.00 | 77.60 |
| **SD** | 7.82 | 5.62 |
| **SEJ** | 1.75 | 1.29 |

[6] An SEJ assumes that panelists are randomly selected and that standard-setting judgments are independent. It is seldom the case that panelists are randomly sampled, and only the first round of judgments may be considered independent. The SEJ, therefore, likely underestimates the uncertainty of passing scores (Tannenbaum & Katz, in press).

In addition to the recommended passing score for each panel, the average passing score across the two panels is provided to help state departments of education determine an appropriate passing score for the Praxis Music: Content Knowledge test. The panels' average passing score recommendation for the Praxis Music: Content Knowledge test is 69.87 (out of a possible 110 raw-score points). The value was rounded to 70 (next highest raw score) to determine the functional recommended passing score. The scaled score associated with 70 raw points is 161.

Table 4 presents the estimated standard error of measurement (SEM) around the recommended passing score[7]. A standard error represents the uncertainty associated with a test score. The scaled score associated with one and two SEMs above and below the recommended passing score are provided.

**Table 4**
***Passing Scores Within 1 and 2 SEMs of the Recommended Passing Score[8]***

| Recommended passing score (SEM) | | Scale score equivalent |
|---|---|---|
| 70 (4.38) | | 161 |
| - 2 SEMs | 62 | 152 |
| -1 SEM | 66 | 157 |
| +1 SEM | 75 | 167 |
| + 2 SEMs | 79 | 172 |

**Summary of Content-specification Judgments**

Panelists judged the extent to which the knowledge reflected by the content specifications was important for entry-level music teachers. Panelists rated the knowledge/skill statements on a four-point scale ranging from *Very Important* to *Not Important*. The panelists' ratings are summarized in Appendix D (see Table D4). All but one of the 44 knowledge/skill statements were judged to be *Very Important* or *Important* by at least three-quarters of the 39 panelists.

[7] The *raw* score SEM value included in this report are updated as data become available. The SEM values listed in each edition of *Understanding Your Praxis Scores* (http://www.ets.org/Media/Tests/PRAXIS/pdf/uyps_web.pdf) are *scaled* score SEM values based on candidate scores on one or more test forms.
[8] The unrounded SEM value is added or subtracted from the unrounded passing score recommendation (see Table 2). The resulting values are rounded up to the next highest whole number and the rounded values are converted to scaled scores.

**Summary of Final Evaluations**

The panelists completed an evaluation form at the conclusion of their standard-setting study. The evaluation form asked the panelists to provide feedback about the quality of the standard-setting implementation and the factors that influenced their decisions.

All panelists *agreed* or *strongly agreed* that they understood the purpose of the study and that the facilitator's instructions and explanations were clear. All panelists *agreed* or *strongly agreed* that they were prepared to make their standard-setting judgments. Across both panels, all panelists *strongly agreed* or *agreed* that the standard-setting process was easy to follow.

All but one of the panelists reported that the definition of the JQC was at least *somewhat influential* in guiding their standard-setting judgments; nearly 80% of panelists indicated the definition was *very influential*. Thirty-six of the 39 panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Nearly three-quarters of the panelists (28 of the 39 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

Across both panels, all of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended; 25 of the 39 panelists were *very comfortable*. Thirty-two of the 39 panelists indicated the recommended passing score was *about right* with six panelists believing the passing score was *too low* and the remaining panelist believing the passing score was *too high*. A summary of the final evaluation results are presented in Appendix D (see Tables D5 and D6).

## Summary

To support the decision-making process for state departments of education with regards to establishing a passing score, or cut score, for the Praxis Music: Content Knowledge (0113) test, research staff from Educational Testing Service designed and conducted a two-panel, multi-state standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level music teachers.

The recommended passing score for each panel, as well as the average passing score across the two panels, are provided to help state departments of education determine an appropriate operational passing score. For the Praxis Music: Content Knowledge test, the recommended passing score[9] is 70 (out of a possible 110 raw-score points). The scaled score associated with a raw score of 70 is 161 (on a 100 - 200 scale).

Panelists judged the extent to which the knowledge and/or skills reflected by the content specifications was important for entry-level music teachers. The favorable judgments of the panelists provided evidence that the content covered by the test is important for beginning practice.

[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. (2010). *The Praxis Series: Music: Content Knowledge (0113) Test at a Glance*. 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.

Tannenbaum, R.J., & Katz, I.R. (forthcoming). Standard setting. In K.F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology*. Washington, DC: American Psychological Association.

# Appendix A
# Panelists' Names & Affiliations

**Praxis Music: Content Knowledge**

| Panelist | Affiliation |
|---|---|
| Olyn Alexander | District of Columbia Public Schools (DC) |
| Melanie Champion | Goodwin Elementary School (CT) |
| Philip Clary | Indian Hill High School (OH) |
| Rachel Cullen | Maple Ave. Elementary School (NH) |
| Christopher Elbing | New Bern High School (NC) |
| Jeanette A. Engelhart | Clark County School Dist. (NV) |
| Camille L. France | Benjamin Banneker High School (DC) |
| Brian Frye | Bancroft Elementary School (DC) |
| Carly Goodman | Clark County School District (NV) |
| Heidi Cohenour Gordon | UAPB (AR) |
| Brandi Hallford | Mansfield R-IV Schools (MO) |
| Jamie M. Hamblin | Casey Co. Middle (KY) |
| Debbie M Headley | Little Rock School District - Baseline Elementary (AR) |
| Daniel S. Hellman | Missouri State University (MO) |
| Joshua Lee Howard | Alleghany County Schools (NC) |
| Ryan Howe | Department of Education (HI) |
| Carolyn Kirio | Kapolei Middle School (HI) |
| Kenneth Martin | Mountain View Middle School (NH) |
| Charles Masters | Central Dauphin High School (PA) |
| Melissa McCabe | Towson University (MD) |
| Brian McCants | Hand Middle School (SC) |
| Brian McGillen | Sudbrook Magnet Middle School (MD) |
| Michael W. Moore | Bob Jones University (SC) |
| David A. Myers | Shelburne Community School (VT) |
| Jennifer Nash | Etna-Dixmont School (ME) |
| Holly Oliver | Plymouth State University (NH) |
| JoAnn Phillips | University of Rhode Island (RI) |
| David A. Rickels | Boise State University (ID) |
| Adam R. Sawyer | Albert D. Lawton Intermediate School (VT) |
| Karin Sehmann | Eastern Kentucky University (KY) |
| Nita Modley Smith | Metro Nashville Public Schools - IT Creswell Arts Magnet (TN) |
| Kathy Stefani | Mountain View School Dist #244 (ID) |
| Nicholas Stokes | Baltimore County Public Schools (MD) |
| Alex Stone | Pine Knot Primary School (KY) |
| Ken Thompson | Bowling Green State University (OH) |
| Linda K. Thompson | Lee University (TN) |
| Jacqueline Townsend | Keystone Local School District (OH) |
| Kristin Wages | Gravette Middle School (AR) |
| Donna Gwyn Wiggins | Winston-Salem State University (NC) |

# Appendix B
# Study Agenda

# AGENDA

## Praxis Music
## Standard Setting Study

Day 1

| | |
|---|---|
| 8:00 – 8:15 | Welcome and Introduction |
| 8:15 – 8:45 | Overview of Standard Setting & the Praxis Music Test |
| 8:45– 10:30 | "Take" the Praxis Music Test |
| 10:30 – 11:00 | Discuss the Praxis Music Test |
| 11:00 – 12:00 | Define the Knowledge/Skills of a JQC |
| 12:00 – 12:45 | Lunch |
| 12:45 – 2:00 | Define the Knowledge/Skills of a JQC (continued) |
| 2:00 – 2:15 | Break |
| 2:15 – 3:00 | Standard Setting Training |
| 3:00 – 5:00 | Round 1 Standard Setting Judgments for Multiple-Choice |
| 5:00 – 5:15 | Collect Materials; End of Day 1 |

# AGENDA

## Praxis Music
## Standard Setting Study

Day 2

| | |
|---|---|
| 9:00 – 9:15 | Overview of Day 2 |
| 9:15 – 9:45 | Standard Setting Training for Constructed-Response |
| 9:45 – 10:15 | Round 1 Standard Setting Judgments for Constructed-Response |
| 10:15 – 10:30 | Break |
| 10:30 – 12:00 | Round 1 Feedback & Round 2 Judgments |
| 12:00 – 12:45 | Lunch |
| 12:45 – 1:45 | Round 1 Feedback & Round 2 Judgments |
| 1:45 – 2:15 | Specification Judgments |
| 2:15 – 2:30 | Feedback on Round 2 Recommended Cut Score |
| 2:30 – 2:45 | Complete Final Evaluation |
| 2:45 – 3:00 | Collect Materials; End of Study |

# Appendix C
# Just Qualified Candidate (JQC) Definition

## Description of a Just Qualified Candidate[10]

NOTE: Where appropriate, the knowledge/skill described refers to materials presented in either aural or written modes

Our JQC …

1. Music History & Theory
    a. Understands period of music history and genres
    b. Understands theory and compositional practices
    c. Is familiar with world musics by regions

2. Performance
    a. Knows instructional techniques for vocal/instrumental rehearsal and performance
    b. Knows warm-up, tuning, and intonation
    c. Knows effective strategies for programming and presentation of performances
    d. Understands accompaniment
    e. Understands how to present proper etiquette

3. Instruction
    a. Knows strategies for instruction, management (classroom and administrative), and assessments for diverse learners and context
    b. Knows local, state, and national standards
    c. Knows basic techniques for improvisation, composition and arranging and how to teach them and other concepts through performance literature
    d. Is aware of approaches for fostering musical expression (performance and creativity)

4. Professional Issues
    a. Is aware of professional responsibilities and situational factors effecting music instruction
    b. Is familiar with careers in musical and advocate music

5. Technology
    a. Understands appropriate use of technology for performance, instruction, recording
    b. Understands ethical and safety issues

6. Instructional Activities
    a. Understands how to select appropriate repertoire
    b. Understands effective pedagogical and instructional techniques
    c. Understands the logistics of concert planning
    d. Is aware that performance problems may have perceptual or production causes
    e. Understands strategies for developing sight-reading skills

[10] Definition appropriate for both Praxis Music: Content Knowledge (0113) and Praxis Music: Content and Instruction (0114) tests.

# Appendix D
# Results for Praxis Music: Content Knowledge

**Table D1**

***Panel Member Demographics (By Panel)***

| | Panel 1 | | Panel 2 | |
|---|---|---|---|---|
| | **N** | **%** | **N** | **%** |
| **Current Position** | | | | |
| Teacher | 14 | 70% | 14 | 74% |
| College Faculty | 6 | 30% | 5 | 26% |
| **Race** | | | | |
| White | 16 | 80% | 15 | 79% |
| Black or African American | 3 | 15% | 2 | 11% |
| Hispanic or Latino | 1 | 5% | 0 | 0% |
| Asian or Asian American | 0 | 0% | 1 | 5% |
| Multiracial | 0 | 0% | 1 | 5% |
| **Gender** | | | | |
| Female | 13 | 65% | 8 | 42% |
| Male | 7 | 35% | 11 | 58% |
| **Which of the following best describes your music education specialty?** | | | | |
| General Music Education | 7 | 35% | 7 | 37% |
| Instructional Music Education | 7 | 35% | 10 | 53% |
| Vocal Music Education | 4 | 20% | 2 | 11% |
| Other | 2 | 10% | 0 | 0% |
| **Are you currently certified as a music teacher?** | | | | |
| Yes | 16 | 80% | 16 | 84% |
| No | 4 | 20% | 3 | 16% |
| **At what K-12 grade level are you currently teaching music** | | | | |
| Elementary (K - 5 or K - 6) | 4 | 20% | 4 | 21% |
| Middle School | 4 | 20% | 3 | 16% |
| Elementary and Middle School | 2 | 10% | 1 | 5% |
| High School | 2 | 10% | 3 | 16% |
| Middle & High School | 2 | 10% | 1 | 5% |
| All Grades | 0 | 0% | 1 | 5% |
| Not currently teaching at the K-12 level | 6 | 30% | 6 | 32% |

**Table D1 (continued)**

***Panel Member Demographics (By Panel)***

| | Panel 1 | | Panel 2 | |
|---|---|---|---|---|
| | **N** | **%** | **N** | **%** |
| **How many years of experience do you have teaching?** | | | | |
| 3 years or less | 2 | 10% | 3 | 16% |
| 4 - 7 years | 5 | 25% | 1 | 5% |
| 8 - 11 years | 1 | 5% | 10 | 53% |
| 12 - 15 years | 3 | 15% | 3 | 16% |
| 16 years or more | 9 | 45% | 2 | 11% |
| **Which best describes the location of your K-12 school?** | | | | |
| Urban | 6 | 30% | 5 | 26% |
| Suburban | 2 | 10% | 6 | 32% |
| Rural | 6 | 30% | 3 | 16% |
| Not currently working in a K-12 school | 6 | 30% | 5 | 26% |
| **If you are college faculty, are you currently involved in the training/preparation of music teachers?** | | | | |
| Yes | 6 | 30% | 5 | 26% |
| No | 0 | 0% | 0 | 0% |
| Not college faculty | 14 | 70% | 14 | 74% |

**Table D2**

**Passing Score Summary by Round of Judgments — Panel 1**

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 65.50 | 66.40 |
| 2 | 76.80 | 77.00 |
| 3 | 70.75 | 70.75 |
| 4 | 72.60 | 72.75 |
| 5 | 81.10 | 81.05 |
| 6 | 68.55 | 70.85 |
| 7 | 70.60 | 70.30 |
| 8 | 66.70 | 67.10 |
| 9 | 76.40 | 77.30 |
| 10 | 79.65 | 79.95 |
| 11 | 84.30 | 84.00 |
| 12 | 68.90 | 69.60 |
| 13 | 75.90 | 78.50 |
| 14 | 72.85 | 73.15 |
| 15 | 59.10 | 60.20 |
| 16 | 49.35 | 49.35 |
| 17 | 69.10 | 71.50 |
| 18 | 70.45 | 72.50 |
| 19 | 75.20 | 78.55 |
| 20 | 69.25 | 68.35 |
| **Average** | 71.15 | 71.96 |
| **Median** | 70.68 | 72.00 |
| **Lowest** | 49.35 | 49.35 |
| **Highest** | 84.30 | 84.00 |
| **SD** | 7.75 | 7.82 |
| **SEJ** | 1.73 | 1.75 |

**Table D3**

**Passing Score Summary by Round of Judgments — Panel 2**

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 59.55 | 61.10 |
| 2 | 65.20 | 69.90 |
| 3 | 64.05 | 65.45 |
| 4 | 67.40 | 69.90 |
| 5 | 70.75 | 71.35 |
| 6 | 57.30 | 58.40 |
| 7 | 70.50 | 73.20 |
| 8 | 68.80 | 68.90 |
| 9 | 65.80 | 65.20 |
| 10 | 58.65 | 67.70 |
| 11 | 55.05 | 56.95 |
| 12 | 71.05 | 72.35 |
| 13 | 76.90 | 77.60 |
| 14 | 63.30 | 64.40 |
| 15 | 73.35 | 71.60 |
| 16 | 73.40 | 74.10 |
| 17 | 71.90 | 72.90 |
| 18 | 60.25 | 61.45 |
| 19 | 64.55 | 65.45 |
| **Average** | 66.20 | 67.78 |
| **Median** | 65.80 | 68.90 |
| **Lowest** | 55.05 | 56.95 |
| **Highest** | 76.90 | 77.60 |
| **SD** | 6.14 | 5.62 |
| **SEJ** | 1.41 | 1.29 |

**Table D4**

***Specification Judgments — Music: Content Knowledge***

| | | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|---|
| | | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| **I.** | **Music History and Theory** | **30** | **77%** | **9** | **23%** | **0** | **0%** | **0** | **0%** |
| | • Understands the history of major developments in musical style and the significant characteristics of important musical styles and historical periods | 26 | 67% | 12 | 31% | 1 | 3% | 0 | 0% |
| | • Is familiar with the style of a variety of world musics and their function in the culture of origin | 12 | 31% | 23 | 59% | 4 | 10% | 0 | 0% |
| | • Understands and analyzes music in aural and written forms and demonstrates aural skills through recognition of melody, harmony, and rhythm | 37 | 95% | 2 | 5% | 0 | 0% | 0 | 0% |
| | • Knows and applies basic music theory concepts when composing, orchestrating, and arranging instrumental and vocal parts | 28 | 72% | 10 | 26% | 1 | 3% | 0 | 0% |
| | • Understands how musical sounds vary | 30 | 77% | 9 | 23% | 0 | 0% | 0 | 0% |
| | • Knows various sources of printed and electronic information on music history and literature | 11 | 28% | 20 | 51% | 8 | 21% | 0 | 0% |
| **II.** | **Performance** | **33** | **85%** | **6** | **15%** | **0** | **0%** | **0** | **0%** |
| | • Demonstrates critical listening skills by identifying errors | 37 | 95% | 2 | 5% | 0 | 0% | 0 | 0% |
| | • Understanding basic conducting techniques | 25 | 64% | 13 | 33% | 1 | 3% | 0 | 0% |
| | • Understands the interpretation of notation and expressive elements for performance in relation to score markings and style periods | 27 | 69% | 11 | 28% | 1 | 3% | 0 | 0% |

**Table D4**

***Specification Judgments — Music: Content Knowledge***

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| • Knows a variety of strategies on how to prepare a musical score for rehearsal and performance | 30 | 77% | 9 | 23% | 0 | 0% | 0 | 0% |
| • Understands basic accompaniment techniques | 5 | 13% | 29 | 74% | 4 | 10% | 1 | 3% |
| • Knows instrumental and choral/vocal warm-up techniques | 30 | 77% | 8 | 21% | 1 | 3% | 0 | 0% |
| • Knows instrumental and choral tuning/intonation technique | 31 | 79% | 7 | 18% | 1 | 3% | 0 | 0% |
| • Understands concert etiquette for performers and audiences and effective techniques for communicating with an audience in a performance setting | 17 | 44% | 20 | 51% | 2 | 5% | 0 | 0% |
| • Understands the practical relationship between acoustics and performance | 8 | 21% | 25 | 64% | 6 | 15% | 0 | 0% |
| **III.A. Instruction, Professional Issues, and Technology: Instruction** | **34** | **87%** | **5** | **13%** | **0** | **0%** | **0** | **0%** |
| • Knows instructional strategies for different class settings | 29 | 74% | 10 | 26% | 0 | 0% | 0 | 0% |
| • Incorporates local, state, and national standards in planning and instruction | 23 | 59% | 14 | 36% | 2 | 5% | 0 | 0% |
| • Understands classroom management techniques | 32 | 82% | 7 | 18% | 0 | 0% | 0 | 0% |
| • Understands how to plan and differentiate instruction | 30 | 77% | 9 | 23% | 0 | 0% | 0 | 0% |
| • Knows strategies to accommodate diverse learning styles and abilities | 30 | 77% | 9 | 23% | 0 | 0% | 0 | 0% |

**Table D4**

***Specification Judgments — Music: Content Knowledge***

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| • Understands how to modify instruction to accommodate student needs | 26 | 67% | 13 | 33% | 0 | 0% | 0 | 0% |
| • Understands a variety of assessment strategies that informs the instructional process | 27 | 69% | 11 | 28% | 1 | 3% | 0 | 0% |
| • Knows and promotes care and maintenance of instruments and materials | 22 | 56% | 12 | 31% | 5 | 13% | 0 | 0% |
| • Knows basic improvisational techniques and how to teach them | 6 | 15% | 28 | 72% | 5 | 13% | 0 | 0% |
| • Knows basic composition and arranging techniques and how to teach them | 10 | 26% | 24 | 62% | 4 | 10% | 1 | 3% |
| • Knows how to teach a variety of musical concepts through performance literature | 29 | 74% | 9 | 23% | 1 | 3% | 0 | 0% |
| • Is aware of approaches for fostering musically expressive experiences | 24 | 62% | 13 | 33% | 2 | 5% | 0 | 0% |
| • Understands how to integrate concepts used in the fine arts and other disciplines in music instruction | 9 | 23% | 25 | 64% | 4 | 10% | 1 | 3% |
| **III.B. Instruction, Professional Issues, and Technology: Professional Issues** | **21** | **54%** | **17** | **44%** | **1** | **3%** | **0** | **0%** |
| • Understands professional ethics and legal issues specific to teaching music | 24 | 62% | 14 | 36% | 1 | 3% | 0 | 0% |
| • Is familiar with the music standards in the National Standards for Arts Education | 23 | 59% | 13 | 33% | 1 | 3% | 2 | 5% |

**Table D4**

***Specification Judgments — Music: Content Knowledge***

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| • Is aware of professional organizations and resources in music education | 11 | 28% | 23 | 59% | 4 | 10% | 1 | 3% |
| • Is aware of current trends and issues in music education | 17 | 44% | 20 | 51% | 2 | 5% | 0 | 0% |
| • Is aware of the major contributions to the history and philosophy of music education and their implications for curriculum | 12 | 31% | 20 | 51% | 7 | 18% | 0 | 0% |
| • Understands philosophical reasons for inclusion of and advocacy for music in the curriculum | 21 | 54% | 16 | 41% | 2 | 5% | 0 | 0% |
| • Recognizes that collaboration with colleagues is important for implementing the curriculum | 22 | 56% | 16 | 41% | 1 | 3% | 0 | 0% |
| • Supports students' learning through two-way communication with parents/guardians | 23 | 59% | 16 | 41% | 0 | 0% | 0 | 0% |
| • Understands basic administrative responsibilities in a music program | 19 | 49% | 17 | 44% | 3 | 8% | 0 | 0% |
| • Is aware of the external influences that affect the music program, curriculum, and student participation | 21 | 54% | 16 | 41% | 2 | 5% | 0 | 0% |
| • Is aware of career opportunities available in music and how to introduce them to students | 5 | 13% | 20 | 51% | 12 | 31% | 2 | 5% |
| • Is aware of strategies for promoting physically healthy performance practices for students and teachers | 18 | 46% | 18 | 46% | 3 | 8% | 0 | 0% |

**Table D4**

***Specification Judgments — Music: Content Knowledge***

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| **III.C. Instruction, Professional Issues, and Technology: Technology** | **23** | **59%** | **15** | **38%** | **1** | **3%** | **0** | **0%** |
| • Understands current technologies used for performance and recording production | 16 | 41% | 20 | 51% | 3 | 8% | 0 | 0% |
| • Is familiar with technology and instructional software and ways to incorporate them in the classroom | 20 | 51% | 18 | 46% | 1 | 3% | 0 | 0% |
| • Demonstrates knowledge of desktop music publishing software for pedagogical purposes | 15 | 38% | 23 | 59% | 1 | 3% | 0 | 0% |
| • Knows the appropriate, ethical, and safe uses for music software and internet technologies | 23 | 59% | 16 | 41% | 0 | 0% | 0 | 0% |

**Table D5**

**Final Evaluation — Panel 1**

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| • I understood the purpose of this study. | 19 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| • The instructions and explanations provided by the facilitators were clear. | 20 | 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. | 19 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| • The explanation of how the recommended passing score is computed was clear. | 17 | 85% | 3 | 15% | 0 | 0% | 0 | 0% |
| • The opportunity for feedback and discussion between rounds was helpful. | 18 | 90% | 2 | 10% | 0 | 0% | 0 | 0% |
| • The process of making the standard setting judgments was easy to follow. | 17 | 85% | 3 | 15% | 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 definition of the JQC | 17 | 85% | 2 | 10% | 1 | 5% |
| • The between-round discussions | 11 | 55% | 8 | 40% | 1 | 5% |
| • The knowledge/skills required to answer each test question | 17 | 85% | 3 | 15% | 0 | 0% |
| • My own professional experience | 14 | 70% | 6 | 30% | 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 for Praxis Music: Content Knowledge (0113)? | 14 | 70% | 6 | 30% | 0 | 0% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *N* | % |
| • Overall, the recommended passing for Praxis Music: Content Knowledge (0113) score is: | 2 | 10% | 17 | 85% | 1 | 5% |

**Table D6**

**Final Evaluation — Panel 2**

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** | ***N*** | **%** |
| • I understood the purpose of this study. | 19 | 100% | 0 | 0% | 0 | 0% | 0 | 0% |
| • The instructions and explanations provided by the facilitators were clear. | 15 | 79% | 4 | 21% | 0 | 0% | 0 | 0% |
| • The training in the standard setting method was adequate to give me the information I needed to complete my assignment. | 16 | 84% | 3 | 16% | 0 | 0% | 0 | 0% |
| • The explanation of how the recommended passing score is computed was clear. | 16 | 84% | 3 | 16% | 0 | 0% | 0 | 0% |
| • The opportunity for feedback and discussion between rounds was helpful. | 16 | 84% | 3 | 16% | 0 | 0% | 0 | 0% |
| • The process of making the standard setting judgments was easy to follow. | 17 | 89% | 2 | 11% | 0 | 0% | 0 | 0% |

**Table 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 definition of the JQC | 14 | 74% | 5 | 26% | 0 | 0% |
| • The between-round discussions | 9 | 47% | 8 | 42% | 2 | 11% |
| • The knowledge/skills required to answer each test question | 15 | 79% | 4 | 21% | 0 | 0% |
| • My own professional experience | 14 | 74% | 5 | 26% | 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 for Praxis Music: Content Knowledge (0113)? | 11 | 58% | 8 | 42% | 0 | 0% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *N* | % |
| • Overall, the recommended passing for Praxis Music: Content Knowledge (0113) score is: | 4 | 21% | 15 | 79% | 0 | 0% |

# Multistate Standard-Setting Technical Report for the

# *Praxis*® Physics (5266)

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

November 2021

Copyright © 2021 by ETS. All rights reserved. ETS, the ETS logo, and PRAXIS are registered trademarks of ETS. 449059763

# Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Physics (5266) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

## Participating States

Panelists representing 11 states were recommended by their respective education agencies. Additionally, the panel included an educator representing Western Governors University (WGU), a nationally-accredited online university. The education agencies and WGU recommended panelists with (a) experience as either teachers of visually impaired students or college faculty who prepare those physics teachers and (b) familiarity with the knowledge and skills required of beginning physics 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* Physics test, the recommended passing score[1] is 56 out of a possible 100 raw-score points. The scale score associated with a raw score of 56 is 145 on a 100–200 scale.

[1] Results from the two panels participating in the study were averaged to produce the recommended passing score. One outlier from Panel 2 was removed from the panel's recommendations before the results from both panels were averaged.

# Introduction

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Physics (5266) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in October 2021. Education agencies[2] recommended panelists with (a) experience as either physics teachers and (b) familiarity with the knowledge and skills required of beginning physics teachers. Eleven states and Western Governors University (WGU), (Table 1) were represented by 19 panelists. (See Appendix A for the names and affiliations of the panelists.)

**Table 1**
***Participating States and Western Governors University and the Number of Panelists***

| | |
|---|---|
| Delaware (2 panelists) | Rhode Island (2 panelists) |
| Idaho (1 panelist) | Tennessee (2 panelists) |
| Indiana (2 panelists) | Utah (2 panelists) |
| Kansas (1 panelist) | Virginia (2 panelists) |
| Maryland (1 panelist) | West Virginia (2 panelists) |
| Mississippi (1 panelist) | Western Governors University (1 panelist) |

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* Physics 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

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

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.

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* Physics 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* Physics 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 need to consider which decision error is more important to minimize.

# Overview of the *Praxis*® Physics Test

The *Praxis*® Physics *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level physics teachers have the knowledge/skills believed necessary for competent professional practice.

The two-hour and 30 minute assessment contains 125 selected-response items[3] covering five content areas: *Nature and Impact of Science and Engineering* (approximately 15 items), *Principles and*

[3] Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate's score.

*Models of Matter and Energy* (approximately 19 items), *Mechanics* (approximately 44 items), *Electricity and Magnetism* (approximately 26 items), and *Waves* (approximately 21 items).[4] Of these 125 items, over 40% also integrate a Science and Enginnering Practice and abou 25% assess content applied to a Task of Teaching Science (ETS, in press). The reporting scale for the *Praxis* Physics test ranges from 100 to 200 scale-score points.

# Processes and Methods

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

Both panels were combined for the beginning of the standard-setting study. The first day began with a welcome and introduction by the meeting facilitators. The lead 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 from both panels were combined as they first took the test and then discussed the content measured as a large group. 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 physics teachers or areas that address content particularly important for entry-level physics teachers.

## Defining the Just-Qualified Candidate

Following the review of the test, panelists described the just-qualified candidate. The *just-qualified candidate description* plays a central role in standard setting (Perie, 2008); the goal of the standard-setting process is to identify the test score that aligns with this description.

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

The panelists created a description of the just-qualified candidate, focusing on the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panelists from both panels were assigned to three smaller groups in order to create a draft description. Then they reconvened and, through whole-group discussion of the three drafts, reached consensus on to determine the final version. This final description of the just-qualified candidate was used by both panels for the remainder of the study.

The description of the just-qualified candidate summarized the panels' 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. A clean, PDF-version of the final description was distributed to panelists to use for the remaining phases of the study (see Appendix C for the just-qualified candidate description).

Given that the two-panel multistate standard-setting study was designed to provide two recommendations for the same performance standard, it was important that panels use a consistent just-qualified candidate description to frame their judgments. Therefore, the panelists from both panels worked together until the just-qualified candidate description was finalized.

## Panelists' Judgments

The standard-setting process for the *Praxis* Physics 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. They then 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. If panelists had any questions, those questions were addressed and retraining was provided, if necessary, before proceeding to the first round of judgments.

Following this first round of judgments (*Round 1*), panel-level summaries and item-level feedback were provided to each panel. For each panel, the panelists' judgments were displayed for each item and summarized across the panelists. Item-level data were highlighted to show when panelists converged in their judgments or diverged in their judgments (i.e., when at least two-thirds of the panelists' judgments were in the same difficulty range).

Each panel 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 for panelists to conform to another's judgment, but for them to understand the different relevant perspectives among the panelists.

During Round 2, each panel discussed their Round 1 judgments and were encouraged by the facilitator to (a) share the rationales for their judgments and (b) 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 their Round 1 judgment. For each panel, the final judgments for the study, therefore, consist of the panelists' 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 and vice versa. The feedback (data), item-level judgments, and resulting discussions for each panel were independent from the other panel.

# Results

## Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panels included 19 educators representing 11 states and WGU. (See Appendix A for a listing of panelists.) Fifteen panelists were teachers, three were college faculty, and one was an instructional coach. Two of the three faculty members' job responsibilities included the training of physics 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)***

| Background Survey Question | Number | Percent |
|---|---|---|
| **What is your current position?** | **N** | **%** |
| Teacher | 15 | 79 |
| College faculty | 3 | 10 |
| Instructional Coach | 1 | 5 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** |
| American Indian or Alaskan Native | 1 | 5 |
| Asian or Asian American | 1 | 5 |
| Black or African American | 2 | 11 |
| White | 14 | 74 |
| Prefer not to respond | 1 | 5 |
| **What is your gender?** | **N** | **%** |
| Female | 8 | 42 |
| Male | 11 | 58 |
| **Are you currently certified to teach physics in your state?** | **N** | **%** |
| Yes | 16 | 84 |
| No | 3 | 16 |
| **Are you currently teaching physics in your state?** | N | % |
| Yes | 17 | 89 |
| No | 2 | 11 |

***(table continues on the next page)***

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

| Background Survey Question | Number | Percent |
|---|---|---|
| **Are you currently supervising or mentoring other physics teachers?** | **N** | **%** |
| Yes | 9 | 47 |
| No | 10 | 53 |
| **At what P–12 grade level are you currently teaching physics?** | **N** | **%** |
| High school (9–12 or 10–12) | 15 | 79 |
| Other | 1 | 5 |
| Not currently teaching at the P–12 level | 3 | 16 |
| **In all, how many years of experience do you have teaching physics?** | **N** | **%** |
| 3 years or less | 1 | 5 |
| 4–7 years | 6 | 32 |
| 8–11 years | 3 | 16 |
| 12–15 years | 2 | 11 |
| 16 years or more | 7 | 37 |
| **Which best describes the location of your P–12 school?** | **N** | **%** |
| Urban | 3 | 16 |
| Suburban | 9 | 47 |
| Rural | 4 | 21 |
| Not currently working at the P–12 level | 3 | 16 |
| **If you are college faculty, are you currently involved in the training/ preparation of physics teacher candidates?** | **N** | **%** |
| Yes | 2 | 11 |
| No | 1 | 5 |
| Not college faculty | 16 | 84 |

## Standard-Setting Judgments

Table 3 summarizes the standard-setting judgments of each panel. The mean represents the panel's passing score recommendation after Round 2 Table 3 includes the summary data with the lowest outlier from Panel 2 removed (High, 2000; see Appendix E). The panelist who provided the lowest score, was believed to be making judgments that would lower the mean, instead of making judgments based on the standard-setting training.

Table 3 also includes the standard deviation 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. 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 E provides the technical notes, which further describe the SEJ.)

**Table 3**

***Summary of Round 2 Standard-setting Judgments by Panel***

| Statistic | Panel 1 | Panel 2 | Panel 2 (1 outlier removed) |
|---|---|---|---|
| Mean | 52.84 | 57.54 | 58.38 |
| Minimum | 47.50 | 50.80 | 54.35 |
| Maximum | 58.75 | 63.50 | 63.50 |
| SD | 3.76 | 3.89 | 3.17 |
| SEJ | 1.19 | 1.30 | 1.12 |

Panelist-level judgments, for Rounds 1 and 2, for each panel are presented in Appendix D (Tables D2 – D5). 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.

The Round 2 mean score is the panel's final recommended passing score. The values were rounded to the next highest whole number to determine the functional recommended passing scores. The panel's passing score recommendation for the *Praxis* Physics test are listed in Table 4, with the outlier removed from Panel 2. The recommendations are out of the 100 raw score points available for the *Praxis* Physics test.

In addition to the recommended passing score for each panel, the mean passing score across both panels is provided to help education agencies determine an appropriate passing score. Table 4 displays those results, also rounded to the next highest whole number to determine the functional recommended passing score.

**Table 4**

***Summary of Round 2 Standard-setting Judgments by Panel***

| | Raw Score | Praxis Score (100 – 200 Scale) |
|---|---|---|
| **Panel 1** | 53 | 141 |
| **Panel 2** | 59 | 150 |
| **Combined** | 56 | 145 |

The conditional standard error of measurement (CSEM) around the recommended passing score of 56 is 4.99 raw points[5]. A standard error represents the uncertainty associated with a test score (See the Technical Notes in Appendix E 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 100 | Praxis Scale Score Equivalent |
|---|---|---|
| RPS - 2 CSEM | 47 | 133 |
| RPS - 1 CSEM | 52 | 140 |
| **RPS** | **56** | **145** |
| RPS +1 CSEM | 61 | 152 |
| RPS +2 CSEM | 66 | 159 |

***Notes.*** CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 4.99 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 their 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, per panel, is presented in Appendix D (Tables D6 – D13).

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* 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.

[5] When all data are included, the average recommended score from both panels is 56 and the CSEM is 4.99 raw score points. These values are the same with the outlier removed.

All panelists reported that the description of the just-qualified candidate was at least *somewhat influential* in guiding their standard-setting judgments. All of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Sixteen of the 19 panelists indicated that their own professional experience was *very influential* in guiding their judgments.

Each panel was presented with their passing score that was calculated during the study, 53 for panel and 58 for panel 2 (out of 100 raw score points). The recommendation to remove the outlier occurred after the meeting. Based on each panel's data, all of the panelists indicated they were at least *somewhat comfortable* with the passing score they recommended. Sixteen of the 19 panelists indicated the recommended passing score was *about right*; one panelist indicated the passing score was *too low* and two indicated it was *too high.*

# Summary

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Physics 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* Physics test, the recommended passing score[6] is 56 out of a possible 100 raw-score points. This recommendation includes the removal of one outlier because the panelist was not believed to be making judgments based on the standard-setting training. The scale score associated with a raw score of 56 is 145 on a 100–200 scale.

[6] 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.

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: Physics (5266)*. 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. (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.

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 and State***

| **Panelist Name** | **Panelists' Affiliation and State Abbreviation** |
|---|---|
| Ranella Anderson | Jim Hill High School (MS) |
| Megan Barnes | Powhatan High School (VA) |
| Brenna Gillman | Western Governors University (WGU) |
| Christian Horner | Westfield High School (IN) |
| Robert Lancaster | Concord High School (DE) |
| Francis Lenox | East Greenwich High School (RI) |
| Stacy Mann | Manchester Valley High School (MD) |
| Milo Maughan | Brighton High School (UT) |
| Wesley Morgan | Nebo School District (UT) |
| Alison Murray | Central Falls High School (RI) |
| Steven Nease | Morristown Hamblen High School West (TN) |
| AdeBanjo Oriade | University of Delaware (DE) |
| Zach Riffle | University High School (WV) |
| Joy Scales | Blacksburg High School (VA) |
| Shayne Seubert | Owyhee High School (ID) |
| Brent Strong | Hauser High School/Ivy Tech (IN) |
| Masakatsu Watanabe | Fort Hays State University (KS) |
| Tamara Westfall | Poca High School (WV) |
| Marcie Williams | Ooltewah High School (TN) |

# Appendix B: Agenda

## *Praxis*® Physics (5266)

## Standard-Setting Study

Day 1 Agenda

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Physics Test

Review the *Praxis* Physics Test

AM Break

Discuss the *Praxis* Physics 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 (small-group drafts) *(continued)*

Collect Materials; End of Day 1

# *Praxis*® Physics (5266)

# Standard-Setting Study

Day 2 Agenda

Overview of Day 2

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

AM Break - Split into two panels

Standard Setting Training in the Modified Angoff Method

Practice Round – Independent Judgments

Lunch

Practice Round –Discussion

Round 1 Standard Setting Judgments

PM Break

Round 1 Standard Setting Judgments *(continued)*

Collect Materials; End of Day 2

# *Praxis*® Physics (5266)

# Standard-Setting Study

Day3 Agenda

Overview of Day 3

Honoraria Presentation

Round 1 Feedback – Panel Summary

Round 1 Feedback and Round 2 Judgments

AM Break

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

Lunch

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[7]**

A just-qualified candidate…

***Nature and Impact of Science and Engineering***

1. Knows how to connect basic principles to real-world applications.
2. Understands experimental design, how the manipulation of variables can test hypotheses, and what can be inferred from the data.
3. Understands units, unit conversions, and significant figures.

***Principles and Models of Matter and Energy***

4. Is familiar with the forms and definitions of temperature, heat, and energy as they relate to thermodynamics
5. Is familiar with conceptual characteristics, processes, and effects of atomic and nuclear interactions.

***Mechanics***

6. Understands the algebraic and graphic nature of linear kinematics.
7. Understands the algebraic and vector nature of 2D motion (including projectile motion and inclined planes).
8. Understands the basics of circular motion.
9. Is familiar with the basic properties of simple harmonic motion, including the motion of springs and pendulums.
10. Is familiar with the basics of rotational motion, including angular momentum, rotational inertia, and torque.
11. Understands applications of Newton's Laws of Motion and gravitation in the analysis of systems and interactions
12. Know and be able to differentiate between impulse and momentum and apply conservation laws in the analysis of 1D systems.
13. Understands how to analyze the mathematical and conceptual interrelation among work, energy, and power.

***Electricity and Magnetism***

14. Knows how to experimentally and theoretically analyze the basics of DC circuits.
15. Is familiar with electric and magnetic forces, fields, and how they are related.
16. Understands the basics of charges and electrostatics.

[7] 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…

***Waves***

17. Understands types of waves and their properties, such as wavelength, amplitude, frequency, and speed.
18. Knows the properties of the electromagnetic spectrum and the relationship between frequency, speed, energy, and wavelength.
19. Is familiar with wave phenomena such as refraction, interference, diffraction, dispersion, scattering and the Doppler Effect.
20. Is familiar with the concepts of simple lenses and mirrors.

***Tasks of Teaching Science / Science and Engineering Practices***

21. Knows inquiry-based instructional methods, scaffolding, and sequencing.
22. Knows how to manipulate formulas to highlight a quantity of interest or proportional relationship.
23. Understands scientific language, discourse, vocabulary, and definitions to explain phenomena, models, and basic representations.

# Appendix D: Panel-Specific Results

**Table D1**

***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **What is your current position?** | **N** | **%** | **N** | **%** |
| Teacher | 8 | 80 | 7 | 78 |
| College faculty | 1 | 10 | 2 | 22 |
| Instructional Coach | 1 | 10 | 0 | 0 |
| **How do you describe yourself (i.e., race/ethnicity)?** | **N** | **%** | **N** | **%** |
| American Indian or Alaskan Native | 0 | 0 | 1 | 11 |
| Asian or Asian American | 1 | 10 | 0 | 0 |
| Black or African American | 1 | 10 | 1 | 11 |
| White | 8 | 80 | 6 | 67 |
| Prefer not to respond | 0 | 0 | 1 | 11 |
| **What is your gender?** | **N** | **%** | **N** | **%** |
| Female | 5 | 50 | 3 | 33 |
| Male | 5 | 50 | 6 | 67 |
| **Are you currently certified to teach physics in your state?** | **N** | **%** | **N** | **%** |
| Yes | 9 | 90 | 7 | 78 |
| No | 1 | 10 | 2 | 22 |
| **Are you currently teaching physics in your state?** | **N** | **%** | **N** | **%** |
| Yes | 9 | 90 | 8 | 89 |
| No | 1 | 10 | 1 | 11 |
| **Are you currently supervising or mentoring other physics teachers?** | **N** | **%** | **N** | **%** |
| Yes | 5 | 50 | 4 | 44 |
| No | 5 | 50 | 5 | 56 |

**Table D1 (*continued from previous page*)**
***Panel Member Demographics per Panel***

| Background Survey Question | Panel 1 Number | Panel 1 Percent | Panel 2 Number | Panel 2 Percent |
|---|---|---|---|---|
| **At what P–12 grade level are you currently teaching physics?** | **N** | **%** | **N** | **%** |
| High school (9–12 or 10–12) | 8 | 80 | 7 | 78 |
| Not currently teaching at the P–12 level | 2 | 20 | 2 | 22 |
| **Including this year, how many years of experience do you have teaching physics?** | **N** | **%** | **N** | **%** |
| 3 years or less | 0 | 0 | 1 | 11 |
| 4–7 years | 5 | 50 | 1 | 11 |
| 8–11 years | 2 | 20 | 1 | 11 |
| 12–15 years | 0 | 0 | 2 | 22 |
| 16 years or more | 3 | 30 | 4 | 44 |
| **Which best describes the location of your P–12 school?** | **N** | **%** | **N** | **%** |
| Urban | 2 | 20 | 1 | 11 |
| Suburban | 5 | 50 | 4 | 44 |
| Rural | 2 | 20 | 2 | 22 |
| Not currently working at the P–12 level | 1 | 10 | 2 | 22 |
| **If you are college faculty, are you currently involved in the training/ preparation of physics teacher candidates?** | **N** | **%** | **N** | **%** |
| Yes | 0 | 0 | 2 | 22 |
| No | 1 | 10 | 0 | 0 |
| Not college faculty | 9 | 90 | 7 | 78 |

**Table D2**

***Panel 1 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 60.70 | 55.35 |
| 2 | 45.40 | 47.50 |
| 3 | 48.30 | 49.90 |
| 4 | 61.75 | 58.75 |
| 5 | 57.50 | 55.10 |
| 6 | 51.60 | 50.80 |
| 7 | 55.35 | 53.15 |
| 8 | 60.00 | 57.00 |
| 9 | 53.15 | 52.70 |
| 10 | 50.60 | 48.10 |

**Table D3**

***Panel 2 Passing Score Summary by Round of Judgments***

| Panelist | Round 1 | Round 2 |
|---|---|---|
| 1 | 56.55 | 56.95 |
| 2 | 53.20 | 50.80 |
| 3 | 54.05 | 54.35 |
| 4 | 58.10 | 58.30 |
| 5 | 62.70 | 62.40 |
| 6 | 60.10 | 59.00 |
| 7 | 55.60 | 55.90 |
| 8 | 56.85 | 56.65 |
| 9 | 64.60 | 63.50 |

**Table D4**

***Summary of Standard-setting Judgments by Panel and by Round (Panel 2 Outlier Removed)***

| Statistic | Panel 1, Round 1 | Panel 1, Round 2 | Panel 2, Round 1 | Panel 2, Round 2 |
|---|---|---|---|---|
| Mean | 54.44 | 52.84 | 57.97 | 58.38 |
| Minimum | 45.40 | 47.50 | 53.20 | 54.35 |
| Maximum | 61.75 | 58.75 | 64.60 | 63.50 |
| SD | 5.55 | 3.76 | 3.84 | 3.17 |
| SEJ | 1.76 | 1.19 | 1.28 | 1.12 |

Note. The judgments from Panelist 2 are a low outlier. It is recommended that the Round 2 judgmnts be removed from the Panel 2 data. The Round 2 summary data from Panel 2 shown in Table D5 includes the outlier.

**Table D5**

***Summary of Panel 2 Standard-setting Judgments (Outlier Included)***

| Statistic | Panel 2, Round 2 |
|---|---|
| Mean | 57.54 |
| Minimum | 50.80 |
| Maximum | 63.50 |
| SD | 3.89 |
| SEJ | 1.30 |

***Table D6: Panel 1 Final Evaluation: 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. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 10 | 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. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 10 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion between rounds was helpful. | 9 | 90 | 1 | 10 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 9 | 90 | 1 | 10 | 0 | 0 | 0 | 0 |

*Table D7: Panel 1 Final Evaluation: 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 | 10 | 100 | 0 | 0 | 0 | 0 |
| The between-round discussions | 7 | 70 | 3 | 30 | 0 | 0 |
| The knowledge/skills required to answer each test item | 8 | 80 | 2 | 20 | 0 | 0 |
| The passing scores of other panel members | 2 | 20 | 8 | 80 | 0 | 0 |
| My own professional experience | 8 | 80 | 2 | 20 | 0 | 0 |

*Table D8: Panel 1 Final Evaluation: 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 % |
|---|---|---|---|---|---|---|---|---|
| Overall, how comfortable are you with the panel's recommended passing score? | 8 | 80 | 2 | 20 | 0 | 0 | 0 | 0 |

*Table D9: Panel 1 Final Evaluation: 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: | 1 | 10 | 8 | 80 | 1 | 10 |

***Table D10: Panel 2 Final Evaluation: 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. | 8 | 89 | 1 | 11 | 0 | 0 | 0 | 0 |
| The instructions and explanations provided by the facilitators were clear. | 8 | 89 | 1 | 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. | 9 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
| The explanation of how the recommended passing score is computed was clear. | 6 | 67 | 3 | 33 | 0 | 0 | 0 | 0 |
| The opportunity for feedback and discussion between rounds was helpful. | 7 | 78 | 2 | 22 | 0 | 0 | 0 | 0 |
| The process of making the standard-setting judgments was easy to follow. | 7 | 78 | 2 | 22 | 0 | 0 | 0 | 0 |

*Table D11: Panel 2 Final Evaluation: 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 | 8 | 89 | 1 | 11 | 0 | 0 |
| The between-round discussions | 3 | 33 | 6 | 67 | 0 | 0 |
| The knowledge/skills required to answer each test item | 8 | 89 | 1 | 11 | 0 | 0 |
| The passing scores of other panel members | 2 | 22 | 6 | 67 | 1 | 11 |
| My own professional experience | 8 | 89 | 1 | 11 | 0 | 0 |

*Table D12: Panel 2 Final Evaluation: 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 % |
|---|---|---|---|---|---|---|---|---|
| Overall, how comfortable are you with the panel's recommended passing score? | 6 | 67 | 3 | 33 | 0 | 0 | 0 | 0 |

*Table D13: Panel 2 Final Evaluation: 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 | 8 | 89 | 1 | 11 |

# Appendix E: 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).

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

## Outlier Analysis

An analysis of the data is conducted per panel. Judgments that are above or below 1.5 times the interquartile range for that panel are identified as outliers (High, 2000). ETS makes recommendations on the removal of *specific* outliers based on the observations of the panel facilitator. The panel facilitator reports whether or not the specified panelist was faithfully participating in the standard-setting process. The decision to accept the panel recommendation with or without the outlier data is solely at the discretion of the state or jurisdiction.

## 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):

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

# Multi-State Standard Setting Report

## Praxis World Languages: German (0183)

## Praxis World Languages: French (0174)

## Praxis World Languages: Spanish (0195)

October 2009

Conducted by

Educational Testing Service

Princeton, New Jersey

Copyright © 2009 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.

## *Executive Summary*

To support the decision-making process for state departments of education with regards to establishing passing scores, or cut scores, for the Praxis World Languages: German, French and Spanish assessments, research staff from Educational Testing Service (ETS) designed and conducted a series of multi-state standard setting studies. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level K-12 German, French and Spanish teachers.

### Recommended Cut Scores

The standard setting studies involved two expert panels for each assessment, comprised of teachers, administrators and college faculty. The recommended cut scores for each panel, as well as the average cut score across the two panels, are provided to help state departments of education determine appropriate cut (or passing) scores.

- For Praxis World Languages: German, the average recommended cut score is 64 (on the raw score metric), which represents 65% of total available 98 raw points (the recommended cut scores for Panels 1 and 2 are 66 and 63, respectively). The scaled score associated with a raw score of 64 on the Praxis German assessment is 163.
- For Praxis World Languages: French, the average recommended cut score is 63 (on the raw score metric), which represents 65% of total available 97 raw points (the recommended cut scores for Panels 1 and 2 are 59 and 66, respectively). The scaled score associated with a raw score of 63 on the Praxis French assessment is 162.
- For Praxis World Languages: Spanish, the recommended cut score is 67 (on the raw score metric), which represents 70% of total available 96 raw points (the recommended cut scores for Panels 1 and 2 are 66 and 69, respectively). The scaled score associated with a raw score of 67 on the Praxis Spanish assessment is 168.

### Summary of Specification Judgments

Panelists judged the extent to which the knowledge and/or skills reflected by the Praxis World Languages assessments content specifications were important for entry-level World Language teachers. For each assessment, all the knowledge/skills statements comprising the test specifications were judged to be *Very Important* or *Important* by a majority of the panelists, providing additional evidence that the content of the Praxis World Languages assessments is important for beginning practice.

## *Introduction*

To support the decision-making process for state departments of education with regards to establishing passing scores, or cut scores, for the Praxis World Languages: German, French and Spanish assessments, research staff from Educational Testing Service (ETS) designed and conducted a series of multi-state standard setting studies. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level K-12 German, French and Spanish teachers. The standard setting studies involved two expert panels for each assessment, comprised of teachers, administrators, and college faculty. Panelists were recommended by departments of education of states that (a) currently use the Praxis Content Knowledge and/or Productive Language Skills assessments or (b) are considering use of the new Praxis World Languages assessments as part of their licensure process.

The design of the multi-state standard setting studies included two, non-overlapping panels to (a) allow each participating state to be represented and (b) replicate the judgment process to strengthen the technical quality of the recommended passing score for each assessment. (See Appendix A for the common agenda used for all panels.)

- German: Two non-overlapping panels with a total of 32 panelists representing 16 states (see Figure 1a) participated.
- French: Two non-overlapping panels with a total of 47 panelists representing 22 states (see Figure 1b) participated.
- Spanish: Two non-overlapping panels with a total of 39 panelists representing 23 states (see Figure 1c) participated.

Figure 1a. Participating States (and number of panelists) for German

| | |
|---|---|
| Alabama (1 panelist) | South Carolina (2 panelists) |
| Delaware (1 panelist) | South Dakota (4 panelists) |
| Kentucky (2 panelists) | Tennessee (4 panelists) |
| Maryland (1 panelist) | Utah (2 panelists) |
| Mississippi (2 panelists) | Wisconsin (1 panelist) |
| North Carolina (2 panelists) | West Virginia (2 panelists) |
| North Dakota (4 panelists) | Wyoming (1 panelist) |
| Pennsylvania (2 panelists) | Nevada (1 panelist) |

NOTE: Alabama, Delaware, Kentucky, Maryland, Wisconsin, Wyoming, and Nevada were represented on only one of the two panels.

Figure 1b. Participating States (and number of panelists) for French

| | |
|---|---|
| Connecticut (2 panelists) | North Dakota (2 panelists) |
| Hawaii (1 panelist) | Pennsylvania (4 panelists) |
| Kentucky (4 panelists) | Rhode Island (1 panelist) |
| Louisiana (3 panelists) | South Carolina (3 panelists) |
| Maine (1 panelist) | South Dakota (1 panelist) |
| Maryland (3 panelists) | Tennessee (3 panelists) |
| Mississippi (4 panelists) | Utah (2 panelists) |
| Missouri (1 panelist) | Vermont (2 panelists) |
| Nevada (2 panelists) | Washington, D.C. (1 panelist) |
| New Hampshire (1 panelist) | West Virginia (2 panelists) |
| North Carolina (2 panelists) | Wisconsin (2 panelists) |

NOTE: Hawaii, Maine, Missouri, New Hampshire, Rhode Island, South Dakota, Vermont, and Washington, D.C., were represented on only one of the two panels.

Figure 1c. Participating States (and number of panelists) for Spanish

| | |
|---|---|
| Alabama (2 panelists) | North Dakota (2 panelists) |
| Delaware (1 panelist) | Ohio (1 panelist) |
| Hawaii (2 panelists) | Pennsylvania (2 panelists) |
| Kentucky (2 panelists) | South Carolina (2 panelists) |
| Louisiana (2 panelists) | South Dakota (2 panelists) |
| Maine (2 panelists) | Tennessee (1 panelist) |
| Maryland (2 panelists) | Utah (1 panelist) |
| Mississippi (2 panelists) | Vermont (3 panelists) |
| Missouri (1 panelist) | Washington, D.C. (1 panelist) |
| Nevada (1 panelist) | West Virginia (3 panelists) |
| New Hampshire (1 panelist) | Wisconsin (1 panelist) |
| North Carolina (2 panelists) | |

NOTE: Delaware, Missouri, Nevada, New Hampshire, Ohio, Pennsylvania, Tennessee, Utah, Washington, D.C., and Wisconsin were represented on only one of the two panels.

The training provided to panelists as well as the study materials were consistent across panels within an assessment with the exception of defining the "just qualified candidate." To assure that both panels for an assessment were using the same frame of reference when making question-level standard setting judgments, the "just qualified candidate" definition developed through a consensus process by the first panel was used as the definition for the second panel. The second panel did complete a thorough review of the definition to allow panelists to internalize the definition. The processes for developing the definition (with Panel 1) and reviewing/internalizing the definition (with Panel 2) are described later, and the "just qualified candidate" definitions are presented in Appendix B.

The panels were convened in July and August 2009 in Princeton, New Jersey. The results for each panel and results combined across panels for each assessment are summarized in the following report. The technical report containing the recommended passing scores for the German, French, and Spanish assessments is provided to each of the represented state departments of education. In each state, the department of education, the state board of education, or a designated educator licensure board is responsible for establishing the final passing scores in accordance with applicable state regulations.

The first national administration of the new Praxis World Languages assessments will occur in fall 2010. The current Praxis Content Knowledge and Productive Language Skills assessments will be phased out, with the last national administration in June 2010 for German and July 2010 for French and Spanish.

## *Praxis World Languages Assessments*

The Praxis World Languages *Test at a Glance* documents (ETS, in press) for the German, French, and Spanish assessments describe the purpose and structure of the assessment. In brief, each assessment measures whether entry-level German, French, or Spanish teachers have the knowledge believed necessary for competent professional practice. A National Advisory Committee of expert practitioners and preparation faculty defined the content of the assessments, and a national survey of the field confirmed the content.

For each of the German, French, and Spanish assessments, the two hour and forty-five minute assessment is divided into four separately timed sections:

- Section I: Listening with Cultural Knowledge (50 minutes) – 36 multiple-choice questions[1]
- Section II: Reading with Cultural Knowledge (50 minutes) – 39 multiple-choice questions[2].
- Section III: Writing (50 minutes) – Three constructed-response questions
- Section IV: Speaking (15 minutes) – Three constructed-response questions.

Candidate scores on the four sections are combined and reported as an overall score; five category scores – Listening, Reading, Cultural Knowledge, Writing, and Speaking – also are reported. The maximum total number of raw points that may be earned on each assessment is 98 for German, 97 for French, and 96 for Spanish. The reporting scales for the Praxis German, French, and Spanish assessments range from 100 to 200 scaled-score points.

[1] For Section I (Listening), 30 of the 36 questions contribute to the candidate's score.
[2] For Section II (Reading), 32 of the 39 questions contribute to the candidate's score for German; 31 of the 39 questions for French; and 30 of the 39 questions for Spanish.

## *Expert Panels*

For each Praxis World Languages assessment, the standard setting study included two expert panels. The various state departments of education recruited panelists to represent a range of professional perspectives. A description of the panels for each assessment is presented below. (See Appendix C for a listing of panelists for each of the six panels.)

### Praxis German Assessment

Panel 1 included 15 teachers, administrators, and college faculty who prepare K-12 German teachers, representing 11 states. In brief, 14 panelists were teachers and one was college faculty. Thirteen panelists were female. Nine panelists indicated they were most fluent in English, and five indicated they were equally fluent in English and German. Fourteen panelists reported being certified German teachers in their states. Approximately half of the panelists had between 4 and 11 years of experience as a K-12 German teacher, and approximately a quarter had 16 or more years of teaching experience.

Panel 2 included 17 teachers, administrators, and college faculty, representing 14 states. In brief, 14 panelists were teachers, one was an administrator, and two were college faculty. Twelve panelists were female. Twelve panelists indicated they were most fluent in English, and five indicated they were equally fluent in English and German. Approximately half of the panelists had 12 or more of experience as a K-12 German teacher, and approximately 20 percent had 3 or fewer years of teaching experience.

A fuller demographic description for the members of the two German panels is presented in Table 1 in Appendix D.

### Praxis French Assessment

Panel 1 included 23 teachers, administrators, and college faculty who prepare K-12 French teachers, representing 18 states. In brief, 15 panelists were teachers, two were administrators, and five were college faculty. Nineteen panelists were White, three were African American, and one was Alaskan Native/American Indian. Seventeen panelists were female. Fourteen panelists indicated they were most fluent in English, and seven indicated they were equally fluent in English and French. Nineteen panelists reported being certified French teachers in their states. Approximately half of the panelists had between 4 and 11 years of experience as a K-12 French teacher, and over a third had 16 or more years of teaching experience.

Panel 2 included 24 teachers, administrators, and college faculty, representing 18 states. In brief, 19 panelists were teachers, two were administrators, and two were college faculty. Nineteen panelists were White, three were African American, and one was Asian American. Eighteen panelists were female. Nineteen panelists indicated they were most fluent in English, and two indicated they were equally fluent in English and French.

Approximately half of the panelists had 16 or more of experience as a K-12 French teacher, and approximately a quarter had 7 or fewer years of teaching experience.

A fuller demographic description for the members of the two French panels is presented in Table 7 in Appendix E.

### Praxis Spanish Assessment

Panel 1 included 18 teachers, administrators, and college faculty who prepare K-12 Spanish teachers, representing 17 states. In brief, 12 panelists were teachers, two were administrators, and four were college faculty. Nine panelists were White, five were Hispanic, three were African American, and one was Asian American. Twelve panelists were female. Thirteen panelists indicated they were most fluent in English, and four indicated they were equally fluent in English and Spanish. Fourteen panelists reported being certified Spanish teachers in their states. Half of the panelists had 16 or more years of experience as a K-12 Spanish teacher, and nearly 40 percent had 11 or fewer years of teaching experience.

Panel 2 included 21 teachers, curriculum specialists, and college faculty, representing 19 states. In brief, 12 panelists were teachers, five were administrators, and four were college faculty. Eight panelists were White, eight were Hispanic, four were African American, and one was Native Hawaiian/Pacific Islander. Sixteen panelists were female. Ten panelists indicated they were most fluent in English, and nine indicated they were equally fluent in English and Spanish. Approximately half of the panelists had 16 or more of experience as a K-12 Spanish teacher, and more than 40 percent had 11 or fewer years of teaching experience.

A fuller demographic description for the members of the two Spanish panels is presented in Table 13 in Appendix F.

## *Process and Method*

The design of the Praxis World Languages assessments standard setting studies included two non-overlapping expert panels for each assessment. As described below, the training provided to panelists and study materials were consistent across panels. Any differences between panels (e.g., defining the "just qualified candidate") are highlighted.

The panelists were sent an e-mail explaining the purpose of the standard-setting study and requesting that they review the test content specifications for the Praxis World Languages assessment (included in the Praxis World Languages *Test at a Glance*, which was attached to the e-mail). The purpose of the review was to familiarize the panelists with the general structure and content of the assessment.

The standard-setting studies began with a welcome and introduction by Drs. Clyde Reese, Patricia Baron, and Wanda Swiggett, ETS researchers in the Center for Validity Research. Dr. Reese, lead facilitator for the studies, then explained how the particular Praxis World Language assessment was developed, provided an overview of standard setting, and presented the agenda for the study.

## Reviewing the Praxis World Languages Assessments

The first activity was for the panelists to "take the test." (Each panelist had signed a nondisclosure form.) The panelists were given approximately two hours to respond to the multiple-choice questions and to sketch responses to the constructed-response questions. The panelists had access to the answer key for the multiple-choice questions and access to the rubrics for the constructed response questions. The purpose of "taking the test" was for the panelists to become familiar with the test format, content, and difficulty.

The panelists then engaged in a discussion of the major content areas being addressed by the assessment; they were also asked to remark on any content areas that they thought would be particularly challenging for entering German, French, or Spanish teachers, and areas that addressed content that would be particularly important for entering teachers.

## Defining the JQC

Following the review of the assessment, panelists internalized the definition of the Just Qualified Candidate (JQC). The JQC is the test taker who has the minimum level of skills believed necessary to be a qualified K-12 German, French, or Spanish teacher. The JQC definition is the operational definition of the cut score. The goal of the standard-setting process is to identify the test score that aligns with this definition of the JQC.

In Panel 1, the panelists were split into smaller groups, and each group was asked to write down their definition of a JQC. Each group referred to Praxis World Languages *Test at a Glance* to guide their definition. Each group posted its definition on chart paper, and a full-panel discussion occurred to reach consensus on a final definition (Appendix B).

In Panel 2, the panelists began with the definition of the JQC developed by the first panel. Given that each multi-state standard setting study was designed to replicate processes and procedures across the two panels for each assessment, it was important that both panels for an assessment use the same JQC definition to frame their judgments. For Panel 2, the panelists reviewed the JQC definition, and any ambiguities were discussed and clarified. The panelists then were split into smaller groups, and each group discussed the behaviors they would expect of the JQC based on the definition and developed performance indicators or "can do" statements based on the definition. The performance indicators were shared across groups and discussed. The purpose of the exercises was to have the panelists internalize the definition.

## Panelists' Judgments

The standard-setting process for the Praxis World Languages assessments was conducted for the overall test, though one standard-setting approach was implemented for Sections I and II (multiple-choice questions) and another approach was implemented for Sections III and IV (constructed-response questions). Each panel's passing score for the assessment is the sum of the interim cut scores recommended by the panelists for each section. These approaches are described next, followed by the results from each standard-setting study. The recommended cut scores for each panel, as well as the average cut score across the two panels, are provided to help state departments of education determine appropriate cut (or passing) scores.

Standard Setting for Sections I and II (Multiple-Choice Questions). A probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) was used for Sections I and II (multiple-choice questions). In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC would answer it 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 a JQC would answer the question correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly.

For each panel, the panelists were asked to approach the judgment process in two stages. First, they reviewed the definition of the JQC and the question and decided if, overall, the question was difficult for the JQC, easy for the JQC, or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rule of thumb to guide their decision:

- difficult questions for a JQC were in the 0 to .30 range;
- easy questions for a JQC were in the .70 to 1 range; and
- moderately difficult/easy questions for a JQC were in the .40 to .60 range.

The second decision was for panelists to decide how they wanted to refine their judgment within the range. For example, if a panelist thought that a question was easy for a JQC, the initial decision located the question in the .70 to 1 range. The second decision was for the panelist to decide if the likelihood of answering it correctly was .70, .80, .90, .95, or 1.0. The two-stage decision-process was implemented to reduce the cognitive load placed on the panelists. The panelists practiced making their standard-setting judgments on the first Listening set (six questions) in Section I.

The panelists engaged in two rounds of judgments. The Round 1 feedback provided to the panel included each panelist's (listed by ID number) recommended cut scores for Sections I and II (as well as cut scores for Sections III and IV) and the panel's average recommended cut score, highest and lowest cut score, and standard

deviation. Following discussion, the panelists' judgments were displayed for each question. The panelists' judgments were summarized by the three general difficulty levels (0 to .30, .40 to .60, and .70 to 1), and the panel's average question judgment was provided. Questions were highlighted to show when panelists converged in their judgments (approximately two-thirds of the panelists located a question in the same difficulty range) or diverged in their judgments. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2).

Other than the definition of the JQC, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

Standard Setting for Sections III and IV (Constructed-Response Questions). An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for Sections III and IV (constructed-response questions). In this approach, for each question, a panelist decides on the assigned score value that would most likely be earned by a JQC. The basic process that each panelist followed was first to review the definition of the JQC and then to review the question and the rubric for that question. The rubric for a question defines holistically the quality of the evidence that would merit a response earning a 3 (*High*), 2 (*Mid-High*), 1 (*Mid-Low*), or 0 (*Low*). During this review, each panelist independently considered the level of knowledge and/or skill required to respond to the question and the features of a response that would earn 3, 2, 1, or 0 points, as defined by the rubric.

A test taker's response to a constructed-response question is independently scored by two raters, and the sum of the raters' scores is the assigned score[3]; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). Each panelist decided on the score most likely to be earned by a JQC from the following possible values: 0, 1, 2, 3, 4, 5, 6. For each of the six constructed-response questions, panelists recorded the score (0 through 6) that a JQC would most likely earn. The panelists practiced making their standard-setting judgments on the first Writing question in Section III.

Consistent with the standard-setting process used for Sections I and II, the panelists engaged in two rounds of judgments for Sections III and IV. After the first round, the judgments of each panelist were summarized and projected for the panel to see and discuss. Each panelist's recommended cut score for Sections III and IV (as well as cut scores for Sections I and II) was displayed as was the panel's average recommended cut score, highest and lowest cut score, and standard deviation. The panelists' judgments also were displayed for each question. The

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

panelists participated in a general discussion of the results. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2).

As with Sections I and II, results from Panel 1 were not shared with the second panel. The question-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

### Judgment of Praxis World Languages Content Specifications

Following the two-round standard setting process, each panel judged the importance of the knowledge and/or skills stated or implied in the assessment content specifications for the job of an entry-level K-12 teacher. The same content specifications were used to develop the German, French, and Spanish assessments. These judgments addressed the perceived content-based validity of the assessment. Judgments were made using a four-point Likert scale — V*ery Important*, I*mportant*, S*lightly Important*, and N*ot Important*. Each panelist independently judged the 21 knowledge/skills statements. (See Appendix G for the common content specifications for the German, French, and Spanish assessments.)

## *Results*

### Initial Evaluation Forms

The panelists completed two initial evaluation forms, once after they were trained in how to make their standard-setting judgments for Sections I and II (multiple-choice questions), and once after they were trained to make their judgments for Sections III and IV (constructed-response questions). The primary information collected from these forms was the panelists indicating if they had received adequate training to make their standard-setting judgments and were ready to proceed. Across all assessments and panels, all panelists indicated that they were prepared to make their judgments.

### Summary of Standard Setting Judgments by Round

A summary of each round of standard-setting judgments for Sections I and II (multiple-choice questions), Sections III and IV (constructed-response questions), and the overall assessment is presented in Appendix D (German), Appendix E (French), and Appendix F (Spanish). The numbers in each table reflect the recommended cut scores — the number of raw points needed to "pass" the section or test — of each panelist for the two rounds. Note that the Praxis World Languages assessments report a single, overall score and that the panels are recommending a single cut score for the combination of Sections I, II, II and IV. The separate "cut scores" for the four sections are intermediate steps in calculating the overall cut score. For each assessment, the panels' average recommended cut score and highest and lowest cut scores are reported, as are the standard deviations (SD) of

panelists' cut scores and the standard errors of judgment (SEJ). The SEJ is one way of estimating the reliability of the judgments. It indicates how likely it would be for other panels of educators similar in make-up, experience, and standard-setting training to the current panels to recommend the same cut score on the same form of the test. A comparable panel's cut score would be within 1 SEJ of the current average cut score 68 percent of the time and within 2 SEJs 95 percent of the time.

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 four of the six panels; the standard deviation increased somewhat between rounds for the first German and Spanish panels.

For each assessment, the Round 2 average score for each section is summed to arrive at each panel's overall recommended cut score (passing score). It should be noted, however, that there are no required minimum section scores that must be obtained in order to pass the German, French, or Spanish assessments. The total test cut score is compensatory, in that as long as the total cut score is met or exceeded, the candidate has passed

## Praxis German Assessment

The panels' cut score recommendations for the Praxis German assessment are 65.71 for Panel 1 and 62.09 for Panel 2 (see Tables 2a and 3a in Appendix D). The values were rounded to the next highest whole number to determine the functional recommended cut scores — 66 for Panel 1 and 63 for Panel 2. The values of 66 and 63 represent approximately 67% and 64%, respectively, of the total available 98 raw points that could be earned on the assessment. The scaled scores associated with 66 and 63 raw points are 165 and 161, respectively.[4]

Tables 4a and 4b (in Appendix D) present the estimated standard errors of measurement (SEM) around the recommended cut scores for each panel. A standard error represents the uncertainty associated with a test score. The scaled scores associated with 1 and 2 SEMs are provided. The standard errors provided are an estimate, given that the Praxis German assessment has not yet been administered.

The recommended cut scores for each panel, as well as the average cut across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score for the Praxis German assessment. The panels' average cut score recommendation for the Praxis German assessment is 63.90. The value was rounded to 64 (next highest raw score) to determine the functional recommended cut score. The value of 64 represents approximately 65% of the total available 98 raw points that could be earned on the assessment.

[4] For reference purposes, if the recommended raw cut score were 65 or 62 points, the scaled score would be 164 or 160, respectively.

The scaled score associated with 64 raw points is 163.[5] Table 4c (in Appendix D) presents the standard error of measurement (SEM) around the recommended cut score combining the information from the two panels.

## Praxis French Assessment

The panels' cut score recommendations for the Praxis French assessment are 58.54 for Panel 1 and 65.84 for Panel 2 (see Tables 8a and 9a in Appendix E). The values were rounded to the next highest whole number to determine the functional recommended cut scores — 59 for Panel 1 and 66 for Panel 2. The values of 59 and 66 represent approximately 61% and 68%, respectively, of the total available 97 raw points that could be earned on the assessment. The scaled scores associated with 59 and 66 raw points are 157 and 166, respectively.[6]

Tables 10a and 10b (in Appendix E) present the estimated standard errors of measurement (SEM) around the recommended cut scores for each panel. A standard error represents the uncertainty associated with a test score. The scaled scores associated with 1 and 2 SEMs are provided. The standard errors provided are an estimate, given that the Praxis French assessment has not yet been administered.

The recommended cut scores for each panel, as well as the average cut across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score for the Praxis French assessment. The panels' average cut score recommendation for the Praxis French assessment is 62.19. The value was rounded to 63 (next highest raw score) to determine the functional recommended cut score. The value of 63 represents approximately 65% of the total available 97 raw points that could be earned on the assessment. The scaled score associated with 63 raw points is 162.[7] Table 10c (in Appendix E) presents the standard error of measurement (SEM) around the recommended cut score combining the information from the two panels.

## Praxis Spanish Assessment

The panels' cut score recommendations for the Praxis Spanish assessment are 65.54 for Panel 1 and 68.02 for Panel 2 (see Tables 14a and 15a in Appendix F). The values were rounded to the next highest whole number to determine the functional recommended cut scores — 66 for Panel 1 and 69 for Panel 2. The values of 66 and 69 represent approximately 69% and 72%, respectively, of the total available 96 raw points that could be earned on the assessment. The scaled scores associated with 66 and 69 raw points are 167 and 171, respectively.[8]

Tables 16a and 16b (in Appendix F) present the estimated standard errors of measurement (SEM) around the recommended cut scores for each panel. A standard error represents the uncertainty associated with a test score.

---

[5] For reference purposes, if the recommended raw cut score was 63 points, the scaled score would be 161.
[6] For reference purposes, if the recommended raw cut score were 58 or 65 points, the scaled score would be 156 or 165, respectively.
[7] For reference purposes, if the recommended raw cut score was 62 points, the scaled score would be 161.
[8] For reference purposes, if the recommended raw cut score were 65 or 68 points, the scaled score would be 166 or 170, respectively.

The scaled scores associated with 1 and 2 SEMs are provided. The standard errors provided are an estimate, given that the Praxis Spanish assessment has not yet been administered.

The recommended cut scores for each panel, as well as the average cut across the two panels, are provided to help state departments of education determine an appropriate cut (or passing) score for the Praxis Spanish assessment. The panels' average cut score recommendation for the Praxis Spanish assessment is 66.78. The value was rounded to 67 (next highest raw score) to determine the functional recommended cut score. The value of 67 represents approximately 70% of the total available 96 raw points that could be earned on the assessment. The scaled score associated with 67 raw points is 168.[9] Table 16c (in Appendix F) presents the standard error of measurement (SEM) around the recommended cut score combining the information from the two panels.

## Summary of Specification Judgments

Panelists judged the extent to which the knowledge and/or skills reflected by the Praxis World Languages assessments content specifications were important for entry-level teachers. Panelists rated the 21 knowledge/skills statements on a four-point scale ranging from *Very Important* to *Not Important*. The panelists' ratings are summarized in Table 5 (in Appendix D) for German, Table 11 (in Appendix E) for French, and Table 17 (in Appendix F) for Spanish.

Across the three assessment, only one knowledge/skills statement — "*Knows how to move beyond literal comprehension in the interpretive mode (listening) by inferring ...*" — was judged to be *Very Important* or *Important* by less than 75% of the panelists for a particular language, German. Two knowledge/skills statements were judged to be *Very Important* or *Important* by less than 90% of the panelists for two languages:

- "*Knows how to communicate orally in the presentational mode (speaking) by delivering oral presentations on familiar literary or cultural topics ...*" for German and Spanish; and
- "*Knows how to contrast syntactical patterns of simple sentences and questions with those of English*" for French and Spanish.

The complete texts of the content specifications are presented in Appendix G.

[9] For reference purposes, if the recommended raw cut score was 66 points, the scaled score would be 167.

## *Summary*

To support the decision-making process for state departments of education with regards to establishing passing scores, or cut scores, for the Praxis World Languages assessments for German, French, and Spanish, research staff from Educational Testing Service (ETS) designed and conducted a series of multi-state standard setting studies. The studies also collected content-related validity evidence to confirm the importance of the content specifications for entry-level K-12 German, French, and Spanish teachers. The standard setting studies involved two expert panels for each assessment, comprised of teachers, administrators, and college faculty.

Standard setting was conducted using a probability-based Angoff approach (for the multiple-choice sections) and an Extended Angoff approach (for the constructed-response sections). Section-level minimum scores were constructed and an overall cut score was computed. The recommended cut scores for each panel, as well as the average cut across the two panels, are provided to help state departments of education determine appropriate cut (or passing) scores.

- For Praxis World Languages: German, the average recommended cut score is 64 (on the raw score metric), which represents 65% of total available 98 raw points (the recommended cut scores for Panels 1 and 2 are 66 and 63, respectively). The scaled score associated with a raw score of 64 on the Praxis German assessment is 163.
- For Praxis World Languages: French, the average recommended cut score is 63 (on the raw score metric), which represents 65% of total available 97 raw points (the recommended cut scores for Panels 1 and 2 are 59 and 66, respectively). The scaled score associated with a raw score of 63 on the Praxis French assessment is 162.
- For Praxis World Languages: Spanish, the recommended cut score is 67 (on the raw score metric), which represents 70% of total available 96 raw points (the recommended cut scores for Panels 1 and 2 are 66 and 69, respectively). The scaled score associated with a raw score of 67 on the Praxis Spanish assessment is 168.

For each of the assessments, both panels confirmed that the knowledge and/or skills stated or implied in the Praxis World Languages assessment content specifications were important for entry-level K-12 teachers. The results of the evaluation surveys (initial and final) from each panels support the quality of the standard-setting implementation.

## *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.

Educational Testing Service. (in press). *German: World Languages: Test at a glance*. Princeton, NJ: Author.

Educational Testing Service. (in press). *French: World Languages: Test at a glance*. Princeton, NJ: Author.

Educational Testing Service. (in press). *Spanish: World Languages: Test at a glance*. Princeton, NJ: Author.

Hambleton, R. K., & Pitoniak, M.J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4 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.

# *APPENDIX A*
# *Common Workshop Agenda*

# AGENDA

## Praxis [Target Language]: World Languages Assessment

## Standard Setting Study

### Day 1

| | |
|---|---|
| **8:00 – 8:15** | **Welcome and Introduction** |
| **8:15 – 8:45** | **Overview of Standard Setting & Workshop Events** |
| **8:45 – 9:15** | **Overview of the Praxis World Languages Assessments** |
| **9:15 – 9:20** | **Break** |
| **9:20 – 11:30** | **"Take" the Praxis [Target Language]: World Languages Assessment** |
| **11:30 – 12:00** | **Discuss the Praxis [Target Language]: World Languages Assessment** |
| **12:00 – 12:15** | **Define the Knowledge/Skills of a JQC** |
| **12:15 – 1:00** | **Lunch** |
| **1:00 – 3:00** | **Define the Knowledge/Skills of a JQC (continued)** |
| **3:00 – 3:15** | **Break** |
| **3:15 – 3:45** | **Standard Setting Training for M-C Questions (Sections I and II)** |
| **3:45 – 5:15** | **Round 1 Standard Setting Judgments for Multiple-Choice** |
| **5:15 – 5:30** | **Collect Materials; End of Day 1** |

# AGENDA

## Praxis [Target Language]: World Languages Assessment

## Standard Setting Study

## Day 2

| | |
|---|---|
| **9:00 – 9:15** | **Questions from Day 1 & Overview of Day 2** |
| **9:15 – 10:00** | **Standard Setting Training for CR Questions (Sections III and IV)** |
| **10:00 – 10:30** | **Round 1 Standard Setting Judgments for Constructed-Response** |
| **10:30 – 10:35** | **Break** |
| **10:35 – 12:00** | **Round 1 Feedback & Round 2 Judgments** |
| **12:00 – 12:45** | **Lunch** |
| **12:45 – 2:15** | **Round 1 Feedback & Round 2 Judgments (continued)** |
| **2:15 – 3:00** | **Specification Judgment** |
| **3:00 – 3:15** | **Feedback on Round 2 Recommended Cut Score** |
| **3:15 – 3:30** | **Complete Final Evaluation** |
| **3:30 – 3:45** | **Collect Materials; End of Study** |

# *APPENDIX B*
# *Just Qualified Candidate (JQC) Definitions*

# *Definition of the Just Qualified Candidate – German*

## Listening, Reading, and Cultural Knowledge

1. Ability to use basic reading strategies, such as word analysis, inference, and context clues, with authentic samples/materials
2. Have a rich, passive German vocabulary which includes high-frequency idioms and grammatical terminology
3. Comprehend a reasonable amount of main ideas, key concepts and some details in authentic samples of paragraph-length discourse
4. In aural and written communication, recognizes various registers and voices to facilitate comprehension
5. Has a basic understanding of syntactical relationships and major verb tenses and moods
6. Can distinguish between phonemes and dipthongs
7. Generally identify significant current, historical, and/or cultural people, places, events, and social structures in German-speaking countries
8. Has a basic understanding of regional differences in language

## Writing and Speaking

1. Ability to adjust pace, intonation, and fluency of delivery
2. Is able to be comprehensible to a native speaker through articulation and pronunciation
3. Can express himself/herself on a variety of concrete and factual topics
4. Has a diverse active vocabulary which allows them to successfully circumlocute, summarize and paraphrase
5. Demonstrates control of mechanics and conventions in writing
6. Demonstrates control of conventions in discourse
7. Is able to adjust writing and speaking for various purposes and audiences
8. Is able to sequence ideas and use conjunctions and transitions to achieve cohesion in writing

# *Definition of the Just Qualified Candidate – French*

## Listening, Reading, and Cultural Knowledge

1. Uses basic reading strategies such as word analysis, inference, and context clues with authentic texts
2. Comprehends a broad French vocabulary including idioms
3. Comprehends (a) main ideas, (b) most key concepts and (c) some details in authentic aural and written communication
4. Recognizes various registers and formal/informal voices to facilitate comprehension in authentic aural and written communication
5. Has an understanding of grammar, including syntax, major verb tenses and moods
6. Has a basic knowledge of French pronunciation
7. Can identify historical or current people, places, events, and social structures in French-speaking countries or regions
8. Has a basic awareness of regional differences in language

## Writing and Speaking

1. Is comprehensible to a native speaker.
2. Can express himself/herself on a variety of concrete and factual topics, including personal opinions
3. Uses a diverse vocabulary to circumlocute, summarize and paraphrase successfully in writing and speaking and engage in conversations
4. Demonstrates basic command of mechanics and conventions in writing
5. Demonstrates control of conventions in speaking
6. Adjusts writing and speaking for various purposes and audiences
7. Sequences ideas to achieve cohesion in writing and speaking

## *Definition of the Just Qualified Candidate – Spanish*

### Listening, Reading, and Cultural Knowledge

1. Uses basic reading strategies such as word analysis, inference, and context clues with authentic texts
2. Comprehends a broad Spanish vocabulary including widely used idiomatic expressions
3. Comprehends (a) main ideas, (b) most subordinate ideas and (c) some details in authentic aural and written communication
4. Comprehends meanings of various registers and formal/informal voice in authentic aural and written communication
5. Has an understanding of grammar, including syntax, verb tenses and moods
6. Has a general knowledge of Spanish pronunciation
7. Can identify historical or current people, places, events, and social structures in Spanish-speaking countries or regions
8. Has a basic awareness of regional differences in language

### Writing and Speaking

1. Is comprehensible to a native speaker.
2. Can express himself/herself on a variety of concrete and factual topics, and express and defend personal opinions
3. Uses a diverse vocabulary to circumlocute, summarize and paraphrase successfully in writing and speaking
4. Appropriately applies mechanics and conventions in writing and speaking
5. Writes and speaks appropriately for various purposes and to varied audiences
6. Sequences ideas to achieve cohesion in writing and speaking

# *APPENDIX C*
# *Panelists' Names & Affiliations*

# German Panel 1

| Panelist | Affiliation |
|---|---|
| Sandra Achenbach | Hardin Valley Academy, Knox County School (TN) |
| Amy L. Bauer | Rapid City Central High School (SD) |
| James H. Bright | Henry Clay High School, Fayette County Public Schools (KY) |
| Mary Ann Crow | Bismarck High School (ND) |
| Stephanie Draheim | Menasha Joint School District (WI) |
| Christi Elkins-Gabbard | Fayette County Schools (KY) |
| VidaJane Haynes | McGavock Comprehensive High School (TN) |
| Brad Martin | Elkins High School (WV) |
| Erin McKeag | Charlotte Mecklenburg Schools (NC) |
| Susan Peterson | T.F. Riggs High School, South Dakota District 32-2 (SD) |
| Colleen Richards | Butler Area School District (PA) |
| Claudia Schoellkopf | Bismarck Public Schools (ND) |
| Wiebke Strehl | University of South Carolina (SC) |
| Shauna Winegar | Mt. Crest High School, Cache County School District (UT) |
| Maga Isabel Wisard | Poplarville Elementary School (MS) |

# German Panel 2

| Panelist | Affiliation |
|---|---|
| Anthony M.DeRosa | Thomas S. Wootton High (MD) |
| Donna M. Evans | Las Vegas Academy/Clark County School District (NV) |
| J. Sarah Floyd | Lexington High School (SC) |
| Sarah Glasser | Wright Jr/Sr High (WY) |
| Melissa Hadorn | Sturgis Brown High School (SD) |
| Arthur D. Holder | Judge Memorial Catholic High School (UT) |
| Diana T. Ihlenfeld | Ohio County Schools (WV) |
| Susanne Lenné Jones | East Carolina University (NC) |
| Elke K. Kuegle | Stevens High School, Rapid City Area Schools (SD) |
| Joy E. Loomis | Newark High School (DE) |
| Joan S. MacDonald | Martin Luther King Magnet (TN) |
| Michelle Mattson | Rhodes College (TN) |
| Cody Mickelson | Jamestown Public School District #1 (ND) |
| Michael C. Netzloff | Bismarck Public Schools (ND) |
| Andrew J. Richards | Fox Chapel Area School District (PA) |
| Dorothée Rosser | Gadsden City High School (AL) |
| Annette Sherrer | Picayune Memorial High School (MS) |

# French Panel 1

| Panelist | Affiliation |
|---|---|
| Anita J. Alkhas | University of Wisconsin-Milwaukee (WI) |
| Pierre C. Baigue | Granite School District (UT) |
| Colette Ballew | Wayne Highlands School District (PA) |
| Claudia V. Bezaka | District of Columbia Public Schools (DC) |
| Paula Summers Calderon | Louisiana State University and A&M College (LA) |
| Cristina Carlotti | East Providence High School (RI) |
| Stephen M. Dubrow | Walter Johnson High School (MD) |
| Nancy Erickson | University of Southern Maine (ME) |
| Gail Fahy | Palo Verde HS Clark County School District (NV) |
| Antoine F.Gnintedem | Sunflower County School District (MS) |
| Melissa Hadorn | Sturgis Brown High School (SD) |
| Sherri K. Harkins | Wicomico County Public Schools (MD) |
| Leanne Hinkle | Bolton High School (TN) |
| Wendy D. Howard | Gaston County School District (NC) |
| Elisabeth Kohl | Council Rock High School –South (PA) |
| William Mann | Clay County High School (WV) |
| Shawn Morrison | College of Charleston (SC) |
| Oscar Niyiragira | Jefferson County Public Schools (KY) |
| Anne Olafson | Minot High School (ND) |
| Amanda Robustelli-Price | Bristol Central High School (CT) |
| Jacquelyn Sergi | South Panola High School (MS) |
| William Thompson | The University of Memphis (TN) |
| Jocelyn A. M. Waddle | Frankfort High School (KY) |

# French Panel 2

| Panelist | Affiliation |
|---|---|
| Lydia Wilson Kohler | George Rogers Clark High School (KY) |
| Robert Desmarais Sullivan | Hattiesburg High School (MS) |
| Denise B. Benskin | Prince Georges County (MD) |
| Crecia C. Swaim | Betsey Ross Arts Magnet School (CT) |
| Jason Bagley | Lexington High School (SC) |
| Mary C. Frye | West Virginia State University (WV) |
| Mary Anne Smith | Pearl City High School (HI) |
| Robert Denis | Las Vegas High School (NV) |
| Nancy Jarchow | Williamstown High School (VT) |
| Madeleine Hooper-Kernen | Missouri State University (MO) |
| Nancy P. Wilson | Mifflin School District (PA) |
| Robert G. Erickson | Brigham Young University (UT) |
| Elizabeth Howe | Hardin Valley Academy (TN) |
| Suzanne Lord Guazzoni | Stone High School (VT) |
| Timothy Wung Kum | Greenville-Weston High School (MS) |
| Stephanie Viator | Cedar Creek School (LA) |
| Wendy C. Mumy | West Craven High School (NC) |
| Jan Hennessey | Dover High School (NH) |
| Tracy Lambert | Lafayette High School (KY) |
| Stephen Keller | A.C. Flora High School (SC) |
| Margaret Schmidt Dess | Shorewood High School (WI) |
| J. Karine Simpson | Central Bucks School District (PA) |
| Linda E. Lassiter | Southern University and A&M College (LA) |
| Valerie Kling | Bismarck High School (ND) |

# Spanish Panel 1

| Panelist | Affiliation |
|---|---|
| Ignacio M. Cariaga | State of Hawaii Public Schools (HI) |
| June C. D. Carter | University of South Carolina Upstate (SC) |
| Eric O. Cintrón | Plymouth State University (NH) |
| Larissa Cuevas | Pass Christian School District (MS) |
| Stephanie Dominguez | Smithville R-II School District (MO) |
| Paul Fallon | East Carolina University (NC) |
| Geoffrey Gillett | Maine School Administrative District 41 (ME) |
| Bridget Suárez Kalmar | Craftsbury Schools (VT) |
| José Labrado | Dawson Springs High School (KY) |
| Mina T. Levenson | Pittsburgh Public Schools (PA) |
| Terri Marlow | Wood County Schools (WV) |
| Belgica Nina-Matos | Delmar School District (DE) |
| Samuel J. Ogdie | Augustana College (SD) |
| Lisa Ramey | North Central Public School (ND) |
| Joyce Richburg | Birmingham City Schools (AL) |
| Ruth E. Smith | University of Louisiana Monroe (LA) |
| Nancy E. Yetter | Baltimore County Public Schools (MD) |
| Thomasina I. White | School District of Philadelphia (PA) |

# Spanish Panel 2

| Panelist | Affiliation |
|---|---|
| Carolyn A. Anderson | Barnwell School District #45 (SC) |
| Isabel Cavour | University of Dayton (OH) |
| Angela Culver Johnson | Madison City Schools (AL) |
| Telece Marbrey | Knox County Schools (TN) |
| Luis M. González-García | Northern Kentucky University (KY) |
| Sharon M. Gracia | Granite School District (UT) |
| Marta C. Gumpert | Southeastern Louisiana University (LA) |
| Andrés V. Hernández | Biloxi Public Schools (MS) |
| David Herren | Union High School (VT) |
| Grace Leavitt | Greely High School / St. Joseph's College (ME) |
| Jennifer Love | Prince George's County Public Schools (MD) |
| Raquel Oxford | University of Wisconsin Milwaukee (WI) |
| Nancy S. Ryan | Berkeley County West Virginia Schools (WV) |
| Ángel T. Tuninetti | West Virginia University (WV) |
| Diane VanDenOever | The University of Sioux Falls (SD) |
| Summer Van Wagnen | Wake County Public School System (NC) |
| Isabel Vázquez-Gil | District of Columbia Public Schools (DC) |
| Nancy Wahineokai | Radford High School (HI) |
| Giovanna Yaranga-Reyes | Burlington School District (VT) |
| James R. Yoder | Clark County School District (NV) |
| Dina Zavala-Petherbridge | Valley City State University (ND) |

# *APPENDIX D*
## *Results for Praxis World Languages: German*

Table 1 Committee Member Demographics — German

| | Panel 1 | | Panel 2 | |
|---|---|---|---|---|
| | N | Percent | N | Percent |
| **Group you are representing** | | | | |
| Teachers | 14 | 93% | 14 | 82% |
| Administrator/Department Head | 0 | 0% | 1 | 6% |
| College Faculty | 1 | 7% | 2 | 12% |
| Other | 0 | 0% | 0 | 0% |
| **Race** | | | | |
| African American or Black | 0 | 0% | 0 | 0% |
| Alaskan Native or American Indian | 0 | 0% | 0 | 0% |
| Asian or Asian American | 0 | 0% | 0 | 0% |
| Native Hawaiian or Other Pacific Islander | 0 | 0% | 0 | 0% |
| White | 14 | 93% | 17 | 100% |
| Hispanic | 1 | 7% | 0 | 0% |
| **Gender** | | | | |
| Female | 13 | 87% | 12 | 71% |
| Male | 2 | 13% | 5 | 29% |
| **In which language are you most fluent?** | | | | |
| English | 9 | 60% | 12 | 71% |
| German | 0 | 0% | 0 | 0% |
| English and German about the same | 5 | 33% | 5 | 29% |
| Other | 1 | 7% | 0 | 0% |
| **Are you certified as a German teacher in your state?** | | | | |
| No | 1 | 7% | 2 | 12% |
| Yes | 14 | 93% | 15 | 88% |
| **Are you currently teaching German in your state?** | | | | |
| No | 1 | 7% | 1 | 6% |
| Yes | 14 | 93% | 16 | 94% |
| **Are you currently mentoring another German teacher?** | | | | |
| No | 14 | 93% | 16 | 94% |
| Yes | 1 | 7% | 1 | 6% |
| **How many years of experience do you have as a German teacher in your state?** | | | | |
| 3 years or less | 1 | 7% | 3 | 18% |
| 4 - 7 years | 4 | 27% | 4 | 24% |
| 8 - 11 years | 4 | 27% | 2 | 12% |
| 12 - 15 years | 2 | 13% | 3 | 18% |
| 16 years or more | 4 | 27% | 5 | 29% |
| **For which education level are you currently teaching German?** | | | | |
| Elementary (K - 5 or K - 6) | 1 | 7% | 0 | 0% |
| Middle School (6 - 8 or 7 - 9) | 2 | 13% | 0 | 0% |
| High School (9 - 12 or 10 - 12) | 9 | 60% | 14 | 82% |
| Middle/High School | 1 | 7% | 1 | 6% |
| Higher Education | 1 | 7% | 2 | 12% |
| Other | 1 | 7% | 0 | 0% |
| **School Setting** | | | | |
| Urban | 8 | 53% | 6 | 35% |
| Suburban | 3 | 20% | 7 | 41% |
| Rural | 4 | 27% | 4 | 24% |

Table 2a Cut score Summary by Round of Judgments — German Panel 1

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 20.74 | 20.51 |
| SD | 1.97 | 2.05 |
| SEJ | 0.51 | 0.53 |
| Highest | 23.70 | 23.20 |
| Lowest | 15.45 | 14.80 |
| Section II: Reading (Max. Raw Score = 32) | | |
| Average | 23.31 | 22.67 |
| SD | 1.43 | 1.41 |
| SEJ | 0.37 | 0.36 |
| Highest | 26.05 | 25.00 |
| Lowest | 20.85 | 19.10 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 11.67 | 11.33 |
| SD | 1.45 | 1.05 |
| SEJ | 0.37 | 0.27 |
| Highest | 15.00 | 13.00 |
| Lowest | 10.00 | 10.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 11.60 | 11.20 |
| SD | 2.29 | 2.01 |
| SEJ | 0.59 | 0.52 |
| Highest | 15.00 | 14.00 |
| Lowest | 7.00 | 7.00 |
| Total (Max. Raw Score = 98) | | |
| Average | 67.32 | 65.71 |
| SD | 5.17 | 5.84 |
| SEJ | 1.34 | 1.51 |
| Highest | 76.90 | 74.20 |
| Lowest | 56.75 | 50.90 |

Table 2b Panelists Cut scores by Round of Judgments — German Panel 1

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 20.40 | 20.60 | 23.80 | 23.30 | 14.00 | 12.00 | 13.00 | 12.00 | 71.20 | 67.90 |
| 2 | 20.40 | 20.05 | 22.40 | 22.00 | 11.00 | 11.00 | 12.00 | 12.00 | 65.80 | 65.05 |
| 3 | 20.75 | 21.05 | 23.30 | 23.40 | 11.00 | 11.00 | 11.00 | 12.00 | 66.05 | 67.45 |
| 4 | 21.40 | 21.90 | 22.95 | 23.85 | 13.00 | 13.00 | 13.00 | 13.00 | 70.35 | 71.75 |
| 5 | 23.70 | 23.20 | 25.35 | 25.00 | 13.00 | 13.00 | 14.00 | 13.00 | 76.05 | 74.20 |
| 6 | 15.45 | 14.80 | 21.30 | 19.10 | 11.00 | 10.00 | 9.00 | 7.00 | 56.75 | 50.90 |
| 7 | 19.50 | 18.50 | 24.80 | 22.80 | 12.00 | 10.00 | 7.00 | 8.00 | 63.30 | 59.30 |
| 8 | 19.80 | 19.90 | 23.00 | 22.90 | 10.00 | 11.00 | 9.00 | 9.00 | 61.80 | 62.80 |
| 9 | 21.85 | 20.90 | 23.60 | 22.20 | 11.00 | 11.00 | 12.00 | 12.00 | 68.45 | 66.10 |
| 10 | 21.10 | 20.50 | 22.45 | 22.20 | 11.00 | 11.00 | 12.00 | 12.00 | 66.55 | 65.70 |
| 11 | 23.00 | 23.20 | 23.90 | 24.10 | 15.00 | 13.00 | 15.00 | 13.00 | 76.90 | 73.30 |
| 12 | 21.60 | 19.40 | 24.00 | 21.85 | 10.00 | 10.00 | 10.00 | 10.00 | 65.60 | 61.25 |
| 13 | 18.75 | 20.45 | 21.95 | 22.65 | 11.00 | 12.00 | 15.00 | 14.00 | 66.70 | 69.10 |
| 14 | 20.85 | 20.70 | 26.05 | 23.65 | 11.00 | 11.00 | 12.00 | 11.00 | 69.90 | 66.35 |
| 15 | 22.60 | 22.45 | 20.85 | 21.05 | 11.00 | 11.00 | 10.00 | 10.00 | 64.45 | 64.50 |

Table 3a Cut score Summary by Round of Judgments — German Panel 2

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 18.48 | 18.09 |
| SD | 2.36 | 2.00 |
| SEJ | 0.57 | 0.48 |
| Highest | 23.55 | 22.65 |
| Lowest | 13.60 | 14.20 |
| Section II: Reading (Max. Raw Score = 32) | | |
| Average | 21.16 | 21.00 |
| SD | 2.28 | 1.86 |
| SEJ | 0.55 | 0.45 |
| Highest | 26.75 | 24.60 |
| Lowest | 17.45 | 17.35 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 12.35 | 12.12 |
| SD | 1.11 | 1.22 |
| SEJ | 0.27 | 0.30 |
| Highest | 14.00 | 14.00 |
| Lowest | 10.00 | 10.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 11.35 | 10.88 |
| SD | 1.62 | 1.11 |
| SEJ | 0.39 | 0.27 |
| Highest | 15.00 | 13.00 |
| Lowest | 8.00 | 9.00 |
| Total (Max. Raw Score = 98) | | |
| Average | 63.34 | 62.09 |
| SD | 4.47 | 4.11 |
| SEJ | 1.08 | 1.00 |
| Highest | 73.30 | 69.25 |
| Lowest | 56.70 | 53.05 |

Table 3b Panelists Cut scores by Round of Judgments — German Panel 2

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 18.50 | 17.70 | 23.80 | 23.10 | 13.00 | 13.00 | 13.00 | 13.00 | 68.30 | 66.80 |
| 2 | 22.20 | 20.10 | 23.40 | 22.20 | 13.00 | 13.00 | 10.00 | 10.00 | 68.60 | 65.30 |
| 3 | 20.25 | 20.05 | 19.90 | 21.10 | 13.00 | 13.00 | 11.00 | 11.00 | 64.15 | 65.15 |
| 4 | 18.65 | 18.25 | 18.70 | 18.75 | 11.00 | 10.00 | 11.00 | 11.00 | 59.35 | 58.00 |
| 5 | 19.90 | 20.20 | 24.00 | 23.70 | 10.00 | 10.00 | 10.00 | 10.00 | 63.90 | 63.90 |
| 6 | 18.30 | 18.10 | 20.10 | 19.20 | 14.00 | 12.00 | 13.00 | 12.00 | 65.40 | 61.30 |
| 7 | 18.95 | 18.05 | 20.75 | 20.45 | 12.00 | 12.00 | 11.00 | 10.00 | 62.70 | 60.50 |
| 8 | 16.90 | 15.70 | 17.45 | 17.35 | 12.00 | 11.00 | 11.00 | 9.00 | 57.35 | 53.05 |
| 9 | 23.55 | 22.65 | 26.75 | 24.60 | 11.00 | 11.00 | 12.00 | 11.00 | 73.30 | 69.25 |
| 10 | 18.40 | 18.20 | 21.40 | 21.10 | 13.00 | 14.00 | 12.00 | 10.00 | 64.80 | 63.30 |
| 11 | 18.20 | 18.00 | 21.15 | 22.65 | 14.00 | 14.00 | 8.00 | 10.00 | 61.35 | 64.65 |
| 12 | 17.75 | 17.75 | 21.55 | 21.25 | 13.00 | 13.00 | 12.00 | 11.00 | 64.30 | 63.00 |
| 13 | 19.20 | 19.20 | 19.80 | 20.50 | 13.00 | 13.00 | 15.00 | 13.00 | 67.00 | 65.70 |
| 14 | 13.60 | 14.20 | 19.00 | 19.30 | 13.00 | 12.00 | 12.00 | 11.00 | 57.60 | 56.50 |
| 15 | 18.55 | 16.45 | 21.50 | 20.60 | 11.00 | 11.00 | 11.00 | 11.00 | 62.05 | 59.05 |
| 16 | 16.10 | 16.95 | 19.90 | 19.95 | 12.00 | 12.00 | 12.00 | 12.00 | 60.00 | 60.90 |
| 17 | 15.20 | 16.00 | 20.50 | 21.20 | 12.00 | 12.00 | 9.00 | 10.00 | 56.70 | 59.20 |

Table 4a Cut scores within 1 and 2 SEMs of the Recommended Cut score — German Panel 1

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 66 (4.50) | | 165 |
| - 2 SEMs | 57 | 153 |
| -1 SEM | 62 | 160 |
| +1 SEM | 71 | 172 |
| + 2 SEMs | 75 | 177 |

Table 4b Cut scores within 1 and 2 SEMs of the Recommended Cut score — German Panel 2

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 63 (4.66) | | 161 |
| - 2 SEMs | 53 | 148 |
| -1 SEM | 58 | 155 |
| +1 SEM | 67 | 166 |
| + 2 SEMs | 72 | 173 |

Table 4c Cut scores within 1 and 2 SEMs of the Recommended Cut score — Combined German Panels

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 64 (4.59) | | 163 |
| - 2 SEMs | 55 | 151 |
| -1 SEM | 60 | 157 |
| +1 SEM | 69 | 169 |
| + 2 SEMs | 74 | 175 |

Note: Consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.

Table 5 Specification Judgments — German (Panels 1 & 2 Judgments Combined)

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| **Language, Linguistics, and Comparison** | | | | | | | | |
| A. Demonstrating Language Proficiency | 27 | 84% | 5 | 16% | 0 | 0% | 0 | 0% |
| Subtopic 1 | 25 | 78% | 7 | 22% | 0 | 0% | 0 | 0% |
| Subtopic 2 | 25 | 78% | 7 | 22% | 0 | 0% | 0 | 0% |
| Subtopic 3 | 14 | 44% | 18 | 56% | 0 | 0% | 0 | 0% |
| Subtopic 4 | 15 | 47% | 16 | 50% | 1 | 3% | 0 | 0% |
| Subtopic 5 | 19 | 59% | 13 | 41% | 0 | 0% | 0 | 0% |
| Subtopic 6 | 21 | 66% | 9 | 28% | 2 | 6% | 0 | 0% |
| Subtopic 7 | 5 | 16% | 19 | 59% | 8 | 25% | 0 | 0% |
| Subtopic 8 | 7 | 22% | 23 | 72% | 2 | 6% | 0 | 0% |
| Subtopic 9 | 27 | 84% | 5 | 16% | 0 | 0% | 0 | 0% |
| Subtopic 10 | 10 | 31% | 22 | 69% | 0 | 0% | 0 | 0% |
| Subtopic 11 | 15 | 47% | 12 | 38% | 4 | 13% | 1 | 3% |
| B. Understanding Linguistics | 18 | 60% | 11 | 37% | 1 | 3% | 0 | 0% |
| Subtopic 1 | 20 | 63% | 12 | 38% | 0 | 0% | 0 | 0% |
| Subtopic 2 | 17 | 53% | 13 | 41% | 2 | 6% | 0 | 0% |
| Subtopic 3 | 17 | 53% | 14 | 44% | 1 | 3% | 0 | 0% |
| Subtopic 4 | 20 | 63% | 12 | 38% | 0 | 0% | 0 | 0% |
| Subtopic 5 | 27 | 84% | 4 | 13% | 1 | 3% | 0 | 0% |
| Subtopic 6 | 19 | 59% | 12 | 38% | 1 | 3% | 0 | 0% |
| C. Comparison of Target Language with English | 13 | 42% | 14 | 45% | 4 | 13% | 0 | 0% |
| Subtopic 1 | 13 | 41% | 16 | 50% | 3 | 9% | 0 | 0% |
| Subtopic 2 | 14 | 44% | 16 | 50% | 2 | 6% | 0 | 0% |
| **Cultures, Literature, Cross-disciplinary Concepts** | | | | | | | | |
| A. Demonstrating Language Proficiency | 16 | 52% | 15 | 48% | 0 | 0% | 0 | 0% |
| Subtopic 1 | 20 | 65% | 10 | 32% | 1 | 3% | 0 | 0% |
| Subtopic 2 | 12 | 39% | 15 | 48% | 4 | 13% | 0 | 0% |

Table 6a Final Evaluation — German Panel 1

| | 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 facilitator were clear. | 14 | 93% | 1 | 7% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods 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 cut scores are computed was clear. | 13 | 87% | 2 | 13% | 0 | 0% | 0 | 0% |
| The opportunity for feedback and discussion between rounds was helpful. | 15 | 100% | 0 | 0% | 0 | 0% | 0 | 0% |
| The process of making the standard setting judgments was easy to follow. | 9 | 60% | 6 | 40% | 0 | 0% | 0 | 0% |

| | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| How influential was each of the following factors in guiding your standard setting judgments? | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 15 | 100% | 0 | 0% | 0 | 0% |
| The between-round discussions | 11 | 73% | 4 | 27% | 0 | 0% |
| The cut scores of other panel members | 4 | 27% | 7 | 47% | 4 | 27% |
| My own professional experience | 10 | 67% | 5 | 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 cut score?[10] | | | | | | | | |

| | Too Low | About Right | Too High |
|---|---|---|---|
| Overall, the panel's recommended cut score is:[10] | | | |

[10] Due to technical problems during the study, panelists were not able to review and judge their comfort level with the overall cut score following Round 2.

Table 6b Final Evaluation — German Panel 2

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| I understood the purpose of this study. | 16 | 94% | 1 | 6% | 0 | 0% | 0 | 0% |
| The instructions and explanations provided by the facilitator were clear. | 13 | 76% | 4 | 24% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods was adequate to give me the information I needed to complete my assignment. | 14 | 82% | 3 | 18% | 0 | 0% | 0 | 0% |
| The explanation of how the recommended cut scores are computed was clear. | 15 | 88% | 2 | 12% | 0 | 0% | 0 | 0% |
| The opportunity for feedback and discussion between rounds was helpful. | 14 | 82% | 3 | 18% | 0 | 0% | 0 | 0% |
| The process of making the standard setting judgments was easy to follow. | 9 | 53% | 7 | 41% | 1 | 6% | 0 | 0% |

| | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| How influential was each of the following factors in guiding your standard setting judgments? | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 14 | 82% | 3 | 18% | 0 | 0% |
| The between-round discussions | 11 | 65% | 2 | 12% | 4 | 24% |
| The cut scores of other panel members | 2 | 12% | 9 | 53% | 6 | 35% |
| My own professional experience | 8 | 47% | 8 | 47% | 1 | 6% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Overall, how comfortable are you with the panel's recommended cut score? | 10 | 59% | 6 | 35% | 1 | 6% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| Overall, the panel's recommended cut score is: | 1 | 6% | 16 | 94% | 0 | 0% |

# *APPENDIX E*
# *Results for Praxis World Languages: French*

Table 7 Committee Member Demographics — French

| | Panel 1 | | Panel 2 | |
|---|---|---|---|---|
| | N | Percent | N | Percent |
| **Group you are representing** | | | | |
| Teachers | 15 | 65% | 19 | 79% |
| Administrator/Department Head | 2 | 9% | 2 | 8% |
| College Faculty | 5 | 22% | 2 | 8% |
| Other | 1 | 4% | 1 | 4% |
| **Race** | | | | |
| African American or Black | 3 | 13% | 3 | 13% |
| Alaskan Native or American Indian | 1 | 4% | 0 | 0% |
| Asian or Asian American | 0 | 0% | 1 | 4% |
| Native Hawaiian or Other Pacific Islander | 0 | 0% | 0 | 0% |
| White | 19 | 83% | 19 | 79% |
| Hispanic | 0 | 0% | 0 | 0% |
| **Gender** | | | | |
| Female | 17 | 74% | 18 | 75% |
| Male | 6 | 26% | 6 | 25% |
| **In which language are you most fluent?** | | | | |
| English | 14 | 61% | 19 | 79% |
| French | 1 | 4% | 3 | 13% |
| English and French about the same | 7 | 30% | 2 | 8% |
| Other | 1 | 4% | 0 | 0% |
| **Are you certified as a French teacher in your state?** | | | | |
| No | 4 | 17% | 4 | 17% |
| Yes | 19 | 83% | 20 | 83% |
| **Are you currently teaching French in your state?** | | | | |
| No | 2 | 9% | 2 | 8% |
| Yes | 21 | 91% | 22 | 92% |
| **Are you currently mentoring another French teacher?** | | | | |
| No | 16 | 70% | 17 | 71% |
| Yes | 7 | 30% | 7 | 29% |
| **How many years of experience do you have as a French teacher in your state?** | | | | |
| 3 years or less | 1 | 4% | 1 | 4% |
| 4 - 7 years | 4 | 17% | 5 | 21% |
| 8 - 11 years | 7 | 30% | 4 | 17% |
| 12 - 15 years | 3 | 13% | 2 | 8% |
| 16 years or more | 8 | 35% | 11 | 46% |
| **For which education level are you currently teaching French?** | | | | |
| Elementary (K - 5 or K - 6) | 2 | 9% | 0 | 0% |
| Middle School (6 - 8 or 7 - 9) | 1 | 4% | 1 | 4% |
| High School (9 - 12 or 10 - 12) | 11 | 48% | 18 | 75% |
| Middle/High School | 2 | 9% | 0 | 0% |
| All Grades (K - 12) | 0 | 0% | 1 | 4% |
| Higher Education | 6 | 26% | 4 | 17% |
| Other | 1 | 4% | 0 | 0% |
| **School Setting** | | | | |
| Urban | 10 | 43% | 9 | 38% |
| Suburban | 6 | 26% | 9 | 38% |
| Rural | 7 | 30% | 6 | 25% |

Table 8a Cut score Summary by Round of Judgments — French Panel 1

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 17.58 | 17.24 |
| SD | 2.24 | 1.90 |
| SEJ | 0.47 | 0.40 |
| Highest | 22.05 | 21.45 |
| Lowest | 13.09 | 14.20 |
| Section II: Reading (Max. Raw Score = 31) | | |
| Average | 21.48 | 21.47 |
| SD | 2.86 | 2.39 |
| SEJ | 0.60 | 0.50 |
| Highest | 28.75 | 27.65 |
| Lowest | 15.00 | 16.20 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 9.78 | 10.26 |
| SD | 1.31 | 1.14 |
| SEJ | 0.27 | 0.24 |
| Highest | 12.00 | 12.00 |
| Lowest | 8.00 | 8.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 9.52 | 9.57 |
| SD | 2.35 | 1.95 |
| SEJ | 0.49 | 0.41 |
| Highest | 16.00 | 16.00 |
| Lowest | 6.00 | 7.00 |
| Total (Max. Raw Score = 97) | | |
| Average | 58.37 | 58.54 |
| SD | 5.33 | 4.56 |
| SEJ | 1.11 | 0.95 |
| Highest | 66.05 | 65.55 |
| Lowest | 45.00 | 48.20 |

Table 8b Panelists Cut scores by Round of Judgments — French Panel 1

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 19.90 | 19.50 | 23.95 | 23.95 | 9.00 | 11.00 | 11.00 | 11.00 | 63.85 | 65.45 |
| 2 | 14.90 | 14.80 | 21.60 | 20.60 | 10.00 | 11.00 | 10.00 | 10.00 | 56.50 | 56.40 |
| 3 | 19.00 | 18.00 | 20.95 | 20.95 | 11.00 | 11.00 | 11.00 | 11.00 | 61.95 | 60.95 |
| 4 | 16.25 | 15.55 | 23.25 | 22.70 | 8.00 | 8.00 | 6.00 | 8.00 | 53.50 | 54.25 |
| 5 | 17.00 | 16.60 | 19.65 | 20.25 | 9.00 | 9.00 | 10.00 | 10.00 | 55.65 | 55.85 |
| 6 | 18.60 | 18.00 | 22.30 | 21.80 | 10.00 | 11.00 | 9.00 | 9.00 | 59.90 | 59.80 |
| 7 | 15.20 | 15.80 | 17.50 | 19.20 | 9.00 | 12.00 | 11.00 | 10.00 | 52.70 | 57.00 |
| 8 | 15.00 | 16.00 | 15.00 | 16.20 | 8.00 | 9.00 | 7.00 | 7.00 | 45.00 | 48.20 |
| 9 | 15.85 | 14.75 | 20.00 | 20.05 | 10.00 | 9.00 | 10.00 | 7.00 | 55.85 | 50.80 |
| 10 | 22.05 | 21.45 | 25.35 | 25.00 | 10.00 | 10.00 | 7.00 | 7.00 | 64.40 | 63.45 |
| 11 | 19.60 | 17.40 | 21.85 | 20.95 | 12.00 | 12.00 | 12.00 | 11.00 | 65.45 | 61.35 |
| 12 | 16.20 | 16.30 | 19.90 | 20.10 | 10.00 | 10.00 | 8.00 | 9.00 | 54.10 | 55.40 |
| 13 | 14.65 | 15.65 | 18.90 | 18.50 | 10.00 | 10.00 | 11.00 | 10.00 | 54.55 | 54.15 |
| 14 | 19.25 | 18.25 | 23.90 | 23.60 | 8.00 | 8.00 | 10.00 | 8.00 | 61.15 | 57.85 |
| 15 | 19.20 | 17.50 | 23.60 | 22.55 | 8.00 | 10.00 | 6.00 | 9.00 | 56.80 | 59.05 |
| 16 | 19.45 | 18.85 | 22.70 | 22.40 | 8.00 | 10.00 | 7.00 | 9.00 | 57.15 | 60.25 |
| 17 | 16.20 | 16.00 | 19.20 | 20.10 | 9.00 | 10.00 | 7.00 | 9.00 | 51.40 | 55.10 |
| 18 | 17.30 | 17.90 | 28.75 | 27.65 | 12.00 | 12.00 | 8.00 | 8.00 | 66.05 | 65.55 |
| 19 | 18.90 | 18.10 | 20.50 | 20.60 | 10.00 | 10.00 | 16.00 | 16.00 | 65.40 | 64.70 |
| 20 | 13.90 | 14.20 | 20.50 | 21.80 | 12.00 | 11.00 | 11.00 | 10.00 | 57.40 | 57.00 |
| 21 | 19.55 | 18.75 | 21.15 | 21.35 | 11.00 | 11.00 | 10.00 | 10.00 | 61.70 | 61.10 |
| 22 | 15.80 | 16.20 | 19.60 | 19.60 | 11.00 | 11.00 | 12.00 | 12.00 | 58.40 | 58.80 |
| 23 | 20.60 | 21.00 | 24.00 | 24.00 | 10.00 | 10.00 | 9.00 | 9.00 | 63.60 | 64.00 |

Table 9a Cut score Summary by Round of Judgments — French Panel 2

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 18.34 | 18.10 |
| SD | 2.46 | 1.96 |
| SEJ | 0.50 | 0.40 |
| Highest | 22.50 | 21.70 |
| Lowest | 14.60 | 15.20 |
| Section II: Reading (Max. Raw Score = 31) | | |
| Average | 22.80 | 23.08 |
| SD | 2.64 | 2.29 |
| SEJ | 0.54 | 0.47 |
| Highest | 27.40 | 27.30 |
| Lowest | 16.40 | 17.40 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 12.00 | 12.67 |
| SD | 1.35 | 1.05 |
| SEJ | 0.28 | 0.21 |
| Highest | 14.00 | 15.00 |
| Lowest | 9.00 | 11.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 11.54 | 12.00 |
| SD | 1.61 | 1.10 |
| SEJ | 0.33 | 0.23 |
| Highest | 14.00 | 14.00 |
| Lowest | 8.00 | 9.00 |
| Total (Max. Raw Score = 97) | | |
| Average | 64.68 | 65.84 |
| SD | 6.03 | 4.68 |
| SEJ | 1.23 | 0.96 |
| Highest | 74.25 | 73.55 |
| Lowest | 52.00 | 56.40 |

Table 9b Panelists Cut scores by Round of Judgments — French Panel 2

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 16.40 | 15.90 | 23.20 | 23.40 | 13.00 | 13.00 | 13.00 | 13.00 | 65.60 | 65.30 |
| 2 | 21.00 | 20.60 | 24.65 | 23.85 | 13.00 | 13.00 | 12.00 | 12.00 | 70.65 | 69.45 |
| 3 | 16.40 | 15.20 | 22.35 | 22.45 | 11.00 | 11.00 | 8.00 | 12.00 | 57.75 | 60.65 |
| 4 | 15.50 | 16.30 | 21.40 | 22.30 | 12.00 | 12.00 | 12.00 | 12.00 | 60.90 | 62.60 |
| 5 | 22.15 | 20.55 | 27.40 | 27.30 | 12.00 | 12.00 | 12.00 | 12.00 | 73.55 | 71.85 |
| 6 | 15.95 | 16.15 | 23.50 | 24.10 | 9.00 | 11.00 | 9.00 | 10.00 | 57.45 | 61.25 |
| 7 | 15.30 | 15.50 | 18.70 | 19.90 | 11.00 | 11.00 | 8.00 | 10.00 | 53.00 | 56.40 |
| 8 | 17.35 | 18.25 | 23.00 | 23.90 | 13.00 | 13.00 | 10.00 | 12.00 | 63.35 | 67.15 |
| 9 | 19.60 | 19.70 | 24.75 | 25.95 | 11.00 | 13.00 | 11.00 | 12.00 | 66.35 | 70.65 |
| 10 | 16.15 | 17.95 | 19.00 | 21.70 | 13.00 | 15.00 | 11.00 | 12.00 | 59.15 | 66.65 |
| 11 | 19.40 | 18.40 | 20.10 | 19.50 | 12.00 | 12.00 | 10.00 | 12.00 | 61.50 | 61.90 |
| 12 | 17.20 | 17.85 | 24.55 | 24.35 | 10.00 | 13.00 | 11.00 | 12.00 | 62.75 | 67.20 |
| 13 | 20.75 | 19.95 | 23.30 | 22.90 | 13.00 | 12.00 | 13.00 | 13.00 | 70.05 | 67.85 |
| 14 | 19.65 | 19.80 | 22.00 | 22.50 | 13.00 | 13.00 | 12.00 | 12.00 | 66.65 | 67.30 |
| 15 | 20.55 | 20.15 | 22.70 | 22.70 | 14.00 | 14.00 | 12.00 | 12.00 | 69.25 | 68.85 |
| 16 | 15.30 | 15.85 | 23.10 | 23.10 | 13.00 | 13.00 | 13.00 | 12.00 | 64.40 | 63.95 |
| 17 | 19.75 | 18.05 | 22.90 | 23.00 | 11.00 | 11.00 | 11.00 | 9.00 | 64.65 | 61.05 |
| 18 | 14.60 | 15.30 | 16.40 | 17.40 | 10.00 | 13.00 | 11.00 | 11.00 | 52.00 | 56.70 |
| 19 | 20.75 | 20.20 | 25.30 | 25.40 | 12.00 | 13.00 | 13.00 | 13.00 | 71.05 | 71.60 |
| 20 | 18.95 | 18.45 | 22.75 | 22.55 | 13.00 | 13.00 | 12.00 | 13.00 | 66.70 | 67.00 |
| 21 | 15.20 | 15.80 | 21.70 | 21.70 | 13.00 | 13.00 | 13.00 | 12.00 | 62.90 | 62.50 |
| 22 | 18.90 | 18.10 | 20.70 | 21.30 | 12.00 | 14.00 | 14.00 | 14.00 | 65.60 | 67.40 |
| 23 | 20.80 | 18.65 | 26.45 | 25.75 | 14.00 | 14.00 | 13.00 | 13.00 | 74.25 | 71.40 |
| 24 | 22.50 | 21.70 | 27.30 | 26.85 | 10.00 | 12.00 | 13.00 | 13.00 | 72.80 | 73.55 |

Table 10a Cut scores within 1 and 2 SEMs of the Recommended Cut score — French Panel 1

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 59 (4.65) | | 157 |
| - 2 SEMs | 50 | 145 |
| -1 SEM | 54 | 150 |
| +1 SEM | 64 | 163 |
| + 2 SEMs | 68 | 169 |

Table 10b Cut scores within 1 and 2 SEMs of the Recommended Cut score — French Panel 2

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 66 (4.54) | | 166 |
| - 2 SEMs | 57 | 154 |
| -1 SEM | 62 | 161 |
| +1 SEM | 71 | 172 |
| + 2 SEMs | 75 | 178 |

Table 10c Cut scores within 1 and 2 SEMs of the Recommended Cut score — Combined French Panels

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 63 (4.61) | | 162 |
| - 2 SEMs | 53 | 149 |
| -1 SEM | 58 | 156 |
| +1 SEM | 67 | 167 |
| + 2 SEMs | 72 | 174 |

Note: Consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.

Table 11 Specification Judgments — French (Panels 1 & 2 Judgments Combined)

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Language, Linguistics, and Comparison | | | | | | | | |
| A. Demonstrating Language Proficiency | 35 | 74% | 12 | 26% | 0 | 0% | 0 | 0 |
| Subtopic 1 | 22 | 47% | 23 | 49% | 2 | 4% | 0 | 0 |
| Subtopic 2 | 36 | 77% | 11 | 23% | 0 | 0% | 0 | 0 |
| Subtopic 3 | 27 | 57% | 19 | 40% | 1 | 2% | 0 | 0% |
| Subtopic 4 | 22 | 47% | 24 | 51% | 1 | 2% | 0 | 0 |
| Subtopic 5 | 31 | 66% | 16 | 34% | 0 | 0% | 0 | 0 |
| Subtopic 6 | 33 | 70% | 12 | 26% | 2 | 4% | 0 | 0 |
| Subtopic 7 | 6 | 13% | 37 | 79% | 4 | 9% | 0 | 0 |
| Subtopic 8 | 11 | 23% | 32 | 68% | 4 | 9% | 0 | 0 |
| Subtopic 9 | 41 | 87% | 6 | 13% | 0 | 0% | 0 | 0 |
| Subtopic 10 | 25 | 53% | 18 | 38% | 3 | 6% | 0 | 0% |
| Subtopic 11 | 19 | 40% | 26 | 55% | 2 | 4% | 0 | 0 |
| B. Understanding Linguistics | 21 | 46% | 23 | 50% | 2 | 4% | 0 | 0 |
| Subtopic 1 | 18 | 40% | 21 | 47% | 6 | 13% | 0 | 0 |
| Subtopic 2 | 24 | 51% | 22 | 47% | 1 | 2% | 0 | 0 |
| Subtopic 3 | 24 | 51% | 21 | 45% | 2 | 4% | 0 | 0 |
| Subtopic 4 | 23 | 50% | 19 | 41% | 4 | 9% | 0 | 0 |
| Subtopic 5 | 24 | 51% | 21 | 45% | 2 | 4% | 0 | 0 |
| Subtopic 6 | 27 | 57% | 18 | 38% | 2 | 4% | 0 | 0 |
| C. Comparison of Target Language with English | 19 | 42% | 20 | 44% | 5 | 11% | 0 | 0% |
| Subtopic 1 | 24 | 51% | 17 | 36% | 5 | 11% | 0 | 0% |
| Subtopic 2 | 15 | 32% | 25 | 53% | 6 | 13% | 0 | 0% |
| Cultures, Literature, Cross-disciplinary Concepts | | | | | | | | |
| A. Demonstrating Language Proficiency | 25 | 53% | 22 | 47% | 0 | 0% | 0 | 0 |
| Subtopic 1 | 23 | 49% | 20 | 43% | 4 | 9% | 0 | 0 |
| Subtopic 2 | 16 | 34% | 28 | 60% | 3 | 6% | 0 | 0 |

Table 12a Final Evaluation — French Panel 1

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| I understood the purpose of this study. | 21 | 91% | 2 | 9% | 0 | 0% | 0 | 0% |
| The instructions and explanations provided by the facilitator were clear. | 18 | 78% | 5 | 22% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods was adequate to give me the information I needed to complete my assignment. | 18 | 78% | 5 | 22% | 0 | 0% | 0 | 0% |
| The explanation of how the recommended cut scores are computed was clear. | 21 | 91% | 2 | 9% | 0 | 0% | 0 | 0% |
| The opportunity for feedback and discussion between rounds was helpful. | 15 | 65% | 6 | 26% | 2 | 9% | 0 | 0% |
| The process of making the standard setting judgments was easy to follow. | 15 | 65% | 8 | 35% | 0 | 0% | 0 | 0% |

| How influential was each of the following factors in guiding your standard setting judgments? | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 20 | 87% | 2 | 9% | 1 | 4% |
| The between-round discussions | 10 | 43% | 12 | 52% | 1 | 4% |
| The cut scores of other panel members | 19 | 83% | 4 | 17% | 0 | 0% |
| My own professional experience | 2 | 9% | 18 | 78% | 3 | 13% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Overall, how comfortable are you with the panel's recommended cut score? | 18 | 78% | 5 | 22% | 0 | 0% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| Overall, the panel's recommended cut score is: | 1 | 4% | 22 | 96% | 0 | 0% |

Table 12b Final Evaluation — French Panel 2

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| I understood the purpose of this study. | 23 | 96% | 1 | 4% | 0 | 0% | 0 | 0% |
| The instructions and explanations provided by the facilitator were clear. | 23 | 96% | 1 | 4% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods was adequate to give me the information I needed to complete my assignment. | 21 | 88% | 3 | 13% | 0 | 0% | 0 | 0% |
| The explanation of how the recommended cut scores are computed was clear. | 19 | 79% | 5 | 21% | 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% |

| How influential was each of the following factors in guiding your standard setting judgments? | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 19 | 79% | 5 | 21% | 0 | 0% |
| The between-round discussions | 15 | 63% | 9 | 38% | 0 | 0% |
| The cut scores of other panel members | 3 | 100% | 0 | 0% | 0 | 0% |
| My own professional experience | 2 | 8% | 16 | 67% | 6 | 25% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Overall, how comfortable are you with the panel's recommended cut score? | 19 | 79% | 4 | 17% | 1 | 4% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| Overall, the panel's recommended cut score is: | 2 | 8% | 22 | 92% | 0 | 0% |

# *APPENDIX F*
## *Results for Praxis World Languages: Spanish*

Table 13 Committee Member Demographics — Spanish

| | Panel 1 | | Panel 2 | |
|---|---|---|---|---|
| | N | Percent | N | Percent |
| **Group you are representing** | | | | |
| Teachers | 12 | 67% | 12 | 57% |
| Administrator/Department Head | 2 | 11% | 5 | 24% |
| College Faculty | 4 | 22% | 4 | 19% |
| Other | 0 | 0% | 0 | 0% |
| **Race** | | | | |
| African American or Black | 3 | 17% | 4 | 19% |
| Alaskan Native or American Indian | 0 | 0% | 0 | 0% |
| Asian or Asian American | 1 | 6% | 0 | 0% |
| Native Hawaiian or Other Pacific Islander | 0 | 0% | 1 | 5% |
| White | 9 | 50% | 8 | 38% |
| Hispanic | 5 | 28% | 8 | 38% |
| **Gender** | | | | |
| Female | 12 | 67% | 16 | 76% |
| Male | 6 | 33% | 5 | 24% |
| **In which language are you most fluent?** | | | | |
| English | 13 | 72% | 10 | 48% |
| Spanish | 1 | 6% | 2 | 10% |
| English and Spanish about the same | 4 | 22% | 9 | 43% |
| Other | 0 | 0% | 0 | 0% |
| **Are you certified as a Spanish teacher in your state?** | | | | |
| No | 4 | 22% | 5 | 24% |
| Yes | 14 | 78% | 16 | 76% |
| **Are you currently teaching Spanish in your state?** | | | | |
| No | 1 | 6% | 4 | 19% |
| Yes | 17 | 94% | 17 | 81% |
| **Are you currently mentoring another Spanish teacher?** | | | | |
| No | 11 | 61% | 10 | 48% |
| Yes | 7 | 39% | 11 | 52% |
| **How many years of experience do you have as a Spanish teacher in your state?** | | | | |
| 3 years or less | 0 | 0% | 0 | 0% |
| 4 - 7 years | 3 | 17% | 1 | 5% |
| 8 - 11 years | 4 | 22% | 8 | 38% |
| 12 - 15 years | 2 | 11% | 2 | 10% |
| 16 years or more | 9 | 50% | 10 | 48% |
| **For which education level are you currently teaching Spanish?** | | | | |
| Elementary (K - 5 or K - 6) | 0 | 0% | 0 | 0% |
| Middle School (6 - 8 or 7 - 9) | 0 | 0% | 1 | 5% |
| High School (9 - 12 or 10 - 12) | 10 | 56% | 12 | 57% |
| Middle/High School | 2 | 11% | 0 | 0% |
| All Grades (K - 12) | 1 | 6% | 2 | 10% |
| Higher Education | 5 | 28% | 6 | 29% |
| **School Setting** | | | | |
| Urban | 9 | 50% | 10 | 48% |
| Suburban | 2 | 11% | 7 | 33% |
| Rural | 7 | 39% | 4 | 19% |

Table 14a Cut score Summary by Round of Judgments — Spanish Panel 1

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 20.18 | 20.23 |
| SD | 2.43 | 2.34 |
| SEJ | 0.57 | 0.55 |
| Highest | 23.95 | 24.05 |
| Lowest | 14.05 | 14.05 |
| Section II: Reading (Max. Raw Score = 31) | | |
| Average | 22.15 | 22.21 |
| SD | 2.56 | 2.63 |
| SEJ | 0.60 | 0.62 |
| Highest | 25.20 | 25.20 |
| Lowest | 15.25 | 14.75 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 11.50 | 11.78 |
| SD | 1.42 | 1.40 |
| SEJ | 0.33 | 0.33 |
| Highest | 14.00 | 14.00 |
| Lowest | 9.00 | 9.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 11.44 | 11.33 |
| SD | 1.50 | 1.24 |
| SEJ | 0.35 | 0.29 |
| Highest | 15.00 | 13.00 |
| Lowest | 9.00 | 9.00 |
| Total (Max. Raw Score = 97) | | |
| Average | 62.27 | 65.54 |
| SD | 5.94 | 5.99 |
| SEJ | 1.40 | 1.41 |
| Highest | 77.65 | 76.25 |
| Lowest | 51.30 | 49.80 |

Table 14b Panelists Cut scores by Round of Judgments — Spanish Panel 1

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 19.80 | 19.80 | 19.30 | 19.30 | 10.00 | 10.00 | 11.00 | 12.00 | 60.10 | 61.10 |
| 2 | 21.20 | 21.50 | 22.85 | 22.85 | 12.00 | 12.00 | 11.00 | 11.00 | 67.05 | 67.35 |
| 3 | 20.25 | 20.35 | 23.75 | 23.75 | 9.00 | 9.00 | 11.00 | 11.00 | 64.00 | 64.10 |
| 4 | 18.40 | 18.70 | 20.85 | 20.85 | 12.00 | 13.00 | 9.00 | 9.00 | 60.25 | 61.55 |
| 5 | 19.35 | 19.95 | 25.05 | 25.05 | 10.00 | 12.00 | 11.00 | 12.00 | 65.40 | 69.00 |
| 6 | 23.45 | 24.05 | 25.20 | 25.20 | 14.00 | 14.00 | 15.00 | 13.00 | 77.65 | 76.25 |
| 7 | 17.65 | 17.65 | 21.60 | 21.60 | 12.00 | 12.00 | 12.00 | 12.00 | 63.25 | 63.25 |
| 8 | 14.05 | 14.05 | 15.25 | 14.75 | 11.00 | 11.00 | 11.00 | 10.00 | 51.30 | 49.80 |
| 9 | 23.50 | 23.60 | 22.95 | 22.95 | 13.00 | 13.00 | 13.00 | 13.00 | 72.45 | 72.55 |
| 10 | 21.55 | 20.85 | 24.05 | 23.55 | 10.00 | 10.00 | 10.00 | 11.00 | 65.60 | 65.40 |
| 11 | 22.75 | 22.35 | 23.85 | 23.75 | 12.00 | 13.00 | 12.00 | 12.00 | 70.60 | 71.10 |
| 12 | 19.65 | 20.55 | 25.05 | 25.00 | 14.00 | 14.00 | 13.00 | 13.00 | 71.70 | 72.55 |
| 13 | 19.40 | 19.60 | 20.50 | 20.20 | 10.00 | 11.00 | 9.00 | 9.00 | 58.90 | 59.80 |
| 14 | 21.20 | 19.80 | 21.15 | 20.65 | 12.00 | 12.00 | 12.00 | 12.00 | 66.35 | 64.45 |
| 15 | 19.80 | 19.80 | 23.70 | 24.20 | 10.00 | 10.00 | 11.00 | 11.00 | 64.50 | 65.00 |
| 16 | 18.50 | 18.70 | 21.30 | 21.80 | 12.00 | 12.00 | 13.00 | 12.00 | 64.80 | 64.50 |
| 17 | 18.80 | 19.45 | 19.05 | 20.05 | 12.00 | 12.00 | 12.00 | 11.00 | 61.85 | 62.50 |
| 18 | 23.95 | 23.35 | 23.20 | 24.20 | 12.00 | 12.00 | 10.00 | 10.00 | 69.15 | 69.55 |

Table 15a Cut score Summary by Round of Judgments — Spanish Panel 2

| | Round 1 | Round 2 |
|---|---|---|
| Section I: Listening (Max. Raw Score = 30) | | |
| Average | 21.76 | 21.47 |
| SD | 2.63 | 2.19 |
| SEJ | 0.57 | 0.48 |
| Highest | 27.00 | 25.45 |
| Lowest | 16.30 | 17.40 |
| Section II: Reading (Max. Raw Score = 31) | | |
| Average | 22.90 | 22.89 |
| SD | 3.27 | 2.74 |
| SEJ | 0.71 | 0.60 |
| Highest | 28.45 | 26.40 |
| Lowest | 15.10 | 16.20 |
| Section III: Writing (Max. Raw Score = 18) | | |
| Average | 12.19 | 12.19 |
| SD | 1.17 | 1.29 |
| SEJ | 0.25 | 0.28 |
| Highest | 14.00 | 15.00 |
| Lowest | 10.00 | 10.00 |
| Section IV: Speaking (Max. Raw Score = 18) | | |
| Average | 11.48 | 11.48 |
| SD | 2.42 | 2.23 |
| SEJ | 0.53 | 0.49 |
| Highest | 14.00 | 15.00 |
| Lowest | 6.00 | 6.00 |
| Total (Max. Raw Score = 97) | | |
| Average | 68.32 | 68.02 |
| SD | 5.97 | 5.91 |
| SEJ | 1.30 | 1.29 |
| Highest | 76.65 | 80.50 |
| Lowest | 51.40 | 54.60 |

Table 15b Panelists Cut scores by Round of Judgments — Spanish Panel 2

| | Section I | | Section II | | Section III | | Section IV | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| Panelist | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 | Rd 1 | Rd 2 |
| 1 | 23.00 | 22.85 | 17.35 | 22.75 | 13.00 | 13.00 | 10.00 | 11.00 | 63.35 | 69.60 |
| 2 | 23.70 | 22.20 | 23.60 | 23.05 | 12.00 | 12.00 | 12.00 | 11.00 | 71.30 | 68.25 |
| 3 | 22.50 | 21.80 | 22.30 | 22.10 | 12.00 | 12.00 | 12.00 | 12.00 | 68.80 | 67.90 |
| 4 | 22.80 | 22.85 | 23.45 | 23.65 | 13.00 | 13.00 | 13.00 | 13.00 | 72.25 | 72.50 |
| 5 | 27.00 | 25.45 | 28.45 | 26.40 | 12.00 | 12.00 | 6.00 | 7.00 | 73.45 | 70.85 |
| 6 | 17.85 | 17.70 | 20.15 | 19.65 | 11.00 | 11.00 | 10.00 | 10.00 | 59.00 | 58.35 |
| 7 | 21.65 | 20.90 | 22.40 | 21.90 | 13.00 | 13.00 | 10.00 | 10.00 | 67.05 | 65.80 |
| 8 | 22.75 | 22.10 | 24.80 | 24.55 | 11.00 | 11.00 | 13.00 | 13.00 | 71.55 | 70.65 |
| 9 | 21.55 | 20.85 | 26.15 | 25.50 | 10.00 | 10.00 | 10.00 | 10.00 | 67.70 | 66.35 |
| 10 | 22.85 | 20.85 | 25.25 | 24.35 | 13.00 | 13.00 | 9.00 | 10.00 | 70.10 | 68.20 |
| 11 | 23.30 | 22.45 | 25.20 | 24.60 | 11.00 | 11.00 | 6.00 | 6.00 | 65.50 | 64.05 |
| 12 | 23.45 | 23.35 | 24.95 | 24.75 | 13.00 | 13.00 | 14.00 | 14.00 | 75.40 | 75.10 |
| 13 | 22.55 | 22.40 | 19.45 | 19.15 | 12.00 | 12.00 | 13.00 | 13.00 | 67.00 | 66.55 |
| 14 | 17.80 | 19.85 | 25.30 | 25.35 | 14.00 | 14.00 | 14.00 | 14.00 | 71.10 | 73.20 |
| 15 | 16.30 | 17.40 | 15.10 | 16.20 | 10.00 | 11.00 | 10.00 | 10.00 | 51.40 | 54.60 |
| 16 | 22.80 | 22.85 | 22.70 | 22.60 | 13.00 | 13.00 | 13.00 | 13.00 | 71.50 | 71.45 |
| 17 | 20.35 | 19.55 | 22.30 | 21.35 | 12.00 | 10.00 | 13.00 | 12.00 | 67.65 | 62.90 |
| 18 | 23.80 | 23.45 | 26.40 | 26.40 | 13.00 | 13.00 | 13.00 | 12.00 | 76.20 | 74.85 |
| 19 | 23.30 | 24.45 | 25.35 | 26.05 | 14.00 | 15.00 | 14.00 | 15.00 | 76.65 | 80.50 |
| 20 | 16.75 | 17.55 | 20.95 | 21.05 | 13.00 | 13.00 | 14.00 | 13.00 | 64.70 | 64.60 |
| 21 | 20.95 | 20.05 | 19.20 | 19.20 | 11.00 | 11.00 | 12.00 | 12.00 | 63.15 | 62.25 |

Table 16a Cut scores within 1 and 2 SEMs of the Recommended Cut score — Spanish Panel 1

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 66 (4.44) | | 167 |
| - 2 SEMs | 57 | 155 |
| -1 SEM | 62 | 162 |
| +1 SEM | 70 | 172 |
| + 2 SEMs | 75 | 179 |

Table 16b Cut scores within 1 and 2 SEMs of the Recommended Cut score — Spanish Panel 2

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 69 (4.33) | | 171 |
| - 2 SEMs | 60 | 159 |
| -1 SEM | 64 | 164 |
| +1 SEM | 73 | 176 |
| + 2 SEMs | 77 | 181 |

Table 16c Cut scores within 1 and 2 SEMs of the Recommended Cut score — Combined Spanish Panels

| Recommended Cut score (SEM) | | Scale Score Equivalent |
|---|---|---|
| 67 (4.38) | | 168 |
| - 2 SEMs | 58 | 156 |
| -1 SEM | 63 | 163 |
| +1 SEM | 72 | 175 |
| + 2 SEMs | 76 | 180 |

Note: Consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.

Table 17 Specification Judgments — Spanish (Panels 1 & 2 Judgments Combined)

| | Very Important | | Important | | Slightly Important | | Not Important | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Language, Linguistics, and Comparison | | | | | | | | |
| A. Demonstrating Language Proficiency | 31 | 84% | 6 | 16% | 0 | 0% | 0 | 0% |
| Subtopic 1 | 26 | 67% | 13 | 33% | 0 | 0% | 0 | 0% |
| Subtopic 2 | 29 | 74% | 10 | 26% | 0 | 0% | 0 | 0% |
| Subtopic 3 | 19 | 49% | 19 | 49% | 1 | 3% | 0 | 0% |
| Subtopic 4 | 20 | 51% | 19 | 49% | 0 | 0% | 0 | 0% |
| Subtopic 5 | 24 | 62% | 15 | 38% | 0 | 0% | 0 | 0% |
| Subtopic 6 | 26 | 67% | 12 | 31% | 1 | 3% | 0 | 0% |
| Subtopic 7 | 10 | 26% | 28 | 72% | 1 | 3% | 0 | 0% |
| Subtopic 8 | 12 | 31% | 27 | 69% | 0 | 0% | 0 | 0% |
| Subtopic 9 | 33 | 85% | 5 | 13% | 1 | 3% | 0 | 0% |
| Subtopic 10 | 18 | 46% | 21 | 54% | 0 | 0% | 0 | 0% |
| Subtopic 11 | 14 | 36% | 21 | 54% | 4 | 10% | 0 | 0% |
| B. Understanding Linguistics | 20 | 53% | 17 | 45% | 0 | 0% | 0 | 0% |
| Subtopic 1 | 20 | 51% | 18 | 46% | 1 | 3% | 0 | 0% |
| Subtopic 2 | 16 | 41% | 20 | 51% | 3 | 8% | 0 | 0% |
| Subtopic 3 | 17 | 44% | 22 | 56% | 0 | 0% | 0 | 0% |
| Subtopic 4 | 23 | 59% | 13 | 33% | 2 | 5% | 0 | 0% |
| Subtopic 5 | 30 | 77% | 8 | 21% | 1 | 3% | 0 | 0% |
| Subtopic 6 | 19 | 49% | 18 | 46% | 2 | 5% | 0 | 0% |
| C. Comparison of Target Language with English | 13 | 35% | 22 | 59% | 2 | 5% | 0 | 0% |
| Subtopic 1 | 15 | 38% | 21 | 54% | 3 | 8% | 0 | 0% |
| Subtopic 2 | 11 | 28% | 24 | 62% | 4 | 10% | 0 | 0% |
| Cultures, Literature, Cross-disciplinary Concepts | | | | | | | | |
| A. Demonstrating Language Proficiency | 15 | 41% | 21 | 57% | 1 | 3% | 0 | 0% |
| Subtopic 1 | 15 | 39% | 20 | 53% | 2 | 5% | 0 | 0% |
| Subtopic 2 | 17 | 45% | 18 | 47% | 3 | 8% | 0 | 0% |

Table 18a Final Evaluation — Spanish Panel 1

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| I understood the purpose of this study. | 18 | 100% | 0 | 0% | 0 | 0% | 0 | 0% |
| The instructions and explanations provided by the facilitator were clear. | 16 | 89% | 2 | 11% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods was adequate to give me the information I needed to complete my assignment. | 15 | 83% | 3 | 17% | 0 | 0% | 0 | 0% |
| The explanation of how the recommended cut scores are computed was clear. | 13 | 72% | 5 | 28% | 0 | 0% | 0 | 0% |
| The opportunity for feedback and discussion between rounds was helpful. | 16 | 94% | 1 | 6% | 0 | 0% | 0 | 0% |
| The process of making the standard setting judgments was easy to follow. | 12 | 67% | 6 | 33% | 0 | 0% | 0 | 0% |

| | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| How influential was each of the following factors in guiding your standard setting judgments? | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 18 | 100% | 0 | 0% | 0 | 0% |
| The between-round discussions | 13 | 72% | 5 | 28% | 0 | 0% |
| The cut scores of other panel members | 6 | 35% | 9 | 53% | 2 | 12% |
| My own professional experience | 16 | 89% | 2 | 11% | 0 | 0% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Overall, how comfortable are you with the panel's recommended cut score? | 11 | 61% | 5 | 28% | 2 | 11% | 0 | 0% |

| | Too Low | | About Right | | Too High | |
|---|---|---|---|---|---|---|
| Overall, the panel's recommended cut score is: | 1 | 6% | 15 | 83% | 2 | 11% |

Table 18b Final Evaluation — Spanish Panel 2

| | Strongly Agree | | Agree | | Disagree | | Strongly Disagree | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| I understood the purpose of this study. | 18 | 90% | 2 | 10% | 0 | 0% | 0 | 0% |
| The instructions and explanations provided by the facilitator were clear. | 19 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| The training in the standard setting methods was adequate to give me the information I needed to complete my assignment. | 19 | 95% | 1 | 5% | 0 | 0% | 0 | 0% |
| The explanation of how the recommended cut scores are computed was clear. | 18 | 90% | 2 | 10% | 0 | 0% | 0 | 0% |
| The opportunity for feedback and discussion between rounds was helpful. | 20 | 100% | 0 | 0% | 0 | 0% | 0 | 0% |
| The process of making the standard setting judgments was easy to follow. | 16 | 80% | 4 | 20% | 0 | 0% | 0 | 0% |

| | Very Influential | | Somewhat Influential | | Not Influential | |
|---|---|---|---|---|---|---|
| How influential was each of the following factors in guiding your standard setting judgments? | N | % | N | % | N | % |
| The definition of the Just Qualified Candidate | 18 | 90% | 2 | 10% | 0 | 0% |
| The between-round discussions | 11 | 55% | 7 | 35% | 2 | 10% |
| The cut scores of other panel members | 0 | 0% | 18 | 90% | 2 | 10% |
| My own professional experience | 16 | 80% | 3 | 15% | 1 | 5% |

| | Very Comfortable | | Somewhat Comfortable | | Somewhat Uncomfortable | | Very Uncomfortable | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Overall, how comfortable are you with the panel's recommended cut score?[11] | | | | | | | | |

| | Too Low | About Right | Too High |
|---|---|---|---|
| Overall, the panel's recommended cut score is:[11] | | | |

[11] Due to technical problems during the study, panelists were not able to review and judge their comfort level with the overall cut score following Round 2.

# *APPENDIX G*
# *Praxis World Languages Content Specifications*

# Language, Linguistics, and Comparisons

## *Demonstrating Language Proficiency*

1. Knows how to communicate in the target language with native speakers unaccustomed to dealing with nonnative speakers, with sufficient accuracy, clarity, and precision to convey the intended message
2. Knows how to communicate in the interpersonal mode (speaking) by participating actively in informal and formal conversations on topics covering home, school, leisure activities, and current events
3. Knows how to communicate in the interpersonal mode (writing) in written exchanges on daily topics
4. Comprehends in the interpretive mode (listening) main ideas and supporting details of audio segments such as news items, short stories, social notices, and reports on familiar topics that deal with factual information
5. Comprehends in the interpretive mode (reading) main ideas and supporting details of printed texts such as news items, short stories, social notices, and reports on familiar topics that deal with factual information
6. Knows how to negotiate meaning in order to sustain an interaction
7. Knows how to move beyond literal comprehension in the interpretive mode (listening) by inferring the meaning of unfamiliar words and phrases in new contexts, inferring and interpreting the author's intent, and offering a personal interpretation of the message
8. Knows how to move beyond literal comprehension in the interpretive mode (reading) by inferring the meaning of unfamiliar words and phrases in new contexts, inferring and interpreting the author's intent, and offering a personal interpretation of the message
9. Understands the gist of normal conversational speech on a variety of topics
10. Knows how to communicate in the presentational mode (writing) by writing routine social correspondence, as well as coherent narratives, descriptions, and summaries about familiar topics of a factual nature in paragraph length in present, past, and future time
11. Knows how to communicate orally in the presentational mode (speaking) by delivering oral presentations on familiar literary or cultural topics and incorporating extra linguistic support to facilitate oral presentations that are extemporaneous or prepared but not read

## *Understanding Linguistics*

1. Understands the rules of the sound system of the target language (i.e., recognizing phonemes and allophones)
2. Recognizes key cohesive devices (conjunctions and adverbs) used in connected discourse
3. Understands high-frequency idiomatic expressions and can infer meaning of words and sentences
4. Knows how to explain the rules that govern the formation of words and sentences in the target language
5. Knows how to exemplify the rules with examples from the target languages, such as the verbal system, pronouns, agreement, word order, interrogatives, both in terms of regularities and irregularities
6. Knows how to identify and use the pragmatic and sociolinguistics conventions and register (formal and informal forms of address)

### *Comparison of Target Language with English*

1. Knows how to identify similarities and differences between the target language and English
2. Knows how to contrast syntactical patterns of simple sentences and questions with those of English

## Cultures, Literatures, Cross-Disciplinary Concepts

### *Demonstrating Cultural Understandings*

1. Knows the three Ps:
    - Perspectives (such as attitudes, ideas, and values)
    - Practices (patterns of behavior and social interaction, such as greetings, turn taking, and rites of passage) and
    - Products (such as tools, foods, law, and music)
2. Recognizes the value and role of authentic literary and cultural texts—such as songs, poems, rhymes and chants, children's books, narrative text, and novels—and usage of those texts to interpret and reflect on the perspectives of the target cultures