



# Indiana Department of Education

Dr. Katie Jenner, Secretary of Education

## MEMORANDUM

**TO:** State Board of Education

**FROM:** Dr. John Keller, Chief Technology Officer  
Risa A. Regnier, Director of Educator Licensing

**DATE:** June 2, 2021

**RE:** Adoption of New and Updated Praxis Tests and Recommended Cut Scores

### Background

In October 2018, the IDOE released an RFP for a new licensure test vendor and program as required by the 2018 General Assembly. In July 2019, in accordance with the results of the RFP and action by the 2019 General Assembly, Educational Testing Service (ETS)--with its nationally available Praxis test program--was adopted by the SBOE to provide educator licensure tests as of September 1, 2021.

Since its adoption in July 2019, ETS and IDOE began working to establish test cut scores for the existing Praxis tests Indiana will use, and the development of several new Praxis tests that will be used in Indiana and made available to other ETS client states. As of December 2020, 42 existing Praxis tests and cut scores had been approved by the board. That level of test and cut score adoption has allowed the DOE to work with the two test vendors to facilitate an earlier transition to many of the Praxis tests than previously planned. The 42 tests and scores already adopted will be available for test administration after July 1, 2021; most of the remaining content tests will be implemented September 1, 2021.

As noted above, in response to Indiana's RFP for licensure tests ETS agreed to develop new Praxis tests in those few required areas for which there is not an existing test, including World Language: Japanese, Music: Instrumental and General Knowledge, and Music: Vocal and General Knowledge. Following the development of these tests a multi-state standard setting (cut score) study was conducted that included Indiana teachers licensed in these content areas as participants. ETS also conducted a score setting study review by Indiana-licensed educators for the existing American Sign Language Proficiency Interview (ASLPI) test used by ETS in its Praxis test suite. The tests and recommended cut scores were posted for public comment for 30 days. There were no public comments submitted.

### Action Needed

Under 511 IAC 15-7-2 the board must vote final approval of cut scores for the licensure tests after 30 days of public comment.

### DOE Recommendation

Final approval of the following Praxis tests and recommended cut scores: 5661 World Language: Japanese (156), 5115 Music: Instrumental and General Knowledge (150), 5116 Music: Vocal and General Knowledge (153), and 0634, the ASLPI (170, or 3+ on the ASLPI 5-point scale).



## Multistate Standard-Setting Technical Report

### ***Praxis® JAPANESE: WORLD LANGUAGE (5661)***

Licensure and Credentialing Research

ETS

Princeton, New Jersey

April 2021

---

# EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis® Japanese: World Language* (5661) test, research staff from Educational Testing Service (ETS) designed and conducted a distance-based multistate standard-setting study.

## PARTICIPATING STATES

Panelists from six states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either Japanese language teachers or college faculty who prepare Japanese language teachers and (b) familiarity with the knowledge and skills required of beginning Japanese language 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 Japanese: World Language* test, the recommended passing score is 59 out of a possible 98 raw-score points. The scale score associated with a raw score of 59 is 156 on a 100–200 scale.

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis® Japanese: World Language* (5661) test, research staff from ETS designed and conducted a distance-based multistate standard-setting study in March 2021. Education agencies<sup>1</sup> recommended panelists with (a) experience as either Japanese language teachers or college faculty who prepare Japanese language teachers and (b) familiarity with the knowledge and skills required of beginning Japanese language teachers. Six states (Table 1) were represented by 10 panelists. (See Appendix A for the names and affiliations of the panelists.)

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

Alaska (1 panelist)	Idaho (2 panelists)
Arkansas (1 panelist)	Indiana (3 panelists)
Iowa (1 panelist)	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, 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 Japanese: World Language* 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 error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the *Praxis Japanese: World Language* 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 Japanese: World Language* test score—is not perfectly reliable. A test score only

<sup>1</sup> States and jurisdictions that currently use *Praxis* tests were invited to participate in the multistate standard-setting study.

*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 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 possesses the required knowledge/skills. The state needs to consider which decision error is more important to minimize.

---

# OVERVIEW OF THE *PRAXIS*<sup>®</sup> JAPANESE: WORLD LANGUAGE TEST

The *Praxis*<sup>®</sup> Japanese: World Language *Study Companion* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures the knowledge, skills, and abilities of examinees who have had preparation in a program for teaching Japanese in grades K–12.

The three-hour assessment contains 75 selected-response items<sup>2,3</sup> and 6 constructed-response items<sup>4</sup> covering four content areas: *Interpretive Listening, including embedded linguistic content* (approximately 30 items), *Interpretive Reading, including embedded linguistic content* (approximately 30 items), *Cultural Knowledge* (approximately 15 items), *Interpersonal and Presentational Writing* (approximately 3 items), and *Interpersonal and Presentational Speaking* (approximately 3 items).<sup>5</sup> The reporting scale for the *Praxis* Japanese: World Language 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 agenda for the panel meeting.

### REVIEWING THE TEST

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

---

<sup>2</sup> Thirteen of the 75 selected-response items are pretest items and do not contribute to a candidate’s score.

<sup>3</sup> Six, non-scored, selected-response items are included as a Listenting Practice section for candidates. They were not included in the standard setting.

<sup>4</sup> One, non-scored, constructed-response item is included as a Writing Practice section for candidates. It was not included in the standard setting.

<sup>5</sup> 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.

The panel created a description of the just qualified candidate —the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, the panel first split into smaller groups to consider the 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 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 *Praxis* Japanese: World Language test includes both dichotomously-scored (selected-response items) and constructed-response items. Panelists received training in two distinct standard-setting approaches: one standard-setting approach for the dichotomously-scored items and another approach for the constructed-response items.

A panel's passing score is the sum of the interim passing scores recommended by the panelists for (a) the dichotomously-scored items and (b) the constructed-response items. As with scoring and reporting, the panelists' judgments for the constructed-response items were weighted such that they contributed 36% of the overall score.

**Dichotomously scored items.** The standard-setting process for the dichotomously-scored items 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.

**Constructed-response items.** An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for the constructed-response items. For this portion of the study, a panelist decided on the assigned score value that would most likely be earned by the just qualified candidate for each constructed-response item. Panelists were asked first to review the definition of the just qualified candidate and then to review the constructed-response item and its rubric. The rubric for a constructed-response item defines (holistically) the quality of the evidence that would merit a response earning a particular score. During this review, each panelist independently considered the level of knowledge/skill required to respond to the constructed-response item and the features of a response that would earn a

particular score, as defined by the rubric. Each panelist decided on the score most likely to be earned by the just qualified candidate from the possible values a test taker can earn.

A test-taker's response to a constructed-response item is independently scored by two raters, and the sum of the raters' scores is the assigned score<sup>6</sup>; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). For their ratings, each panelist decided on the score most likely to be earned by a just qualified candidate from the following possible values: 0, 1, 2, 3, 4, 5, or 6. For each of the constructed-response item, panelists recorded the score (0 through 6) that a just qualified candidate would most likely earn.

After the training, panelists made practice judgments and discussed those judgments and their rationale. All panelists completed a post-training survey to confirm that they had received adequate training and felt prepared to continue; the standard-setting process continued only if all panelists confirmed their readiness.

**Multiple Rounds.** Following this first round of judgments (*Round 1*), item-level feedback was provided to the panel. The panelists' judgments were displayed for each item and summarized across panelists. For dichotomously-scored items, items were highlighted to show when panelists converged in their judgments (at least two-thirds of the panelists located an item in the same difficulty range) or diverged in their judgments.

The panelists discussed their item-level judgments. These discussions helped panelists maintain a shared understanding of the knowledge/skills of the 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.

---

<sup>6</sup> If the two raters' scores differ by more than one point (non-adjacent), the Chief Reader for that item assigns the score, which is then doubled.

---

## RESULTS

### EXPERT PANELS

Table 2 presents a summary of the panelists' demographic information. The panel included 10 educators representing six states. (See Appendix A for a listing of panelists.) Six panelists were teachers, two were college faculty, one was an administrator or department head, and one was an Upward Bound instructor. One of the three faculty members' job responsibilities included the training of Japanese language teachers.

**Table 2**  
*Panel Member Demographics*

	<i>N</i>	<i>%</i>
<b>Current position</b>		
Teacher	6	60
Administrator/Department head	1	10
College faculty	2	20
Upward Bound Instructor	1	10
<b>Race</b>		
Asian or Asian American	5	50
White	4	40
Prefer not to respond	1	10
<b>Gender</b>		
Female	4	40
Male	6	60
<b>Are you currently certified to teach Japanese in your state?</b>		
Yes	8	80
No	2	20
<b>Are you currently teaching Japanese in your state?</b>		
Yes	9	90
No	1	10
<b>Are you currently supervising or mentoring other Japanese language teachers?</b>		
Yes	2	20
No	8	80

**Table 2 (continued)****Panel Member Demographics**

	<i>N</i>	<i>%</i>
<b>At what K–12 grade level are you currently teaching Japanese?</b>		
Middle school (6–8 or 7–9)	1	10
Middle and High School	1	10
High school (9–12 or 10–12)	6	60
Not currently teaching at the P–12 level	2	20
<b>Including this year, how many years of experience do you have teaching Japanese?</b>		
3 years or less	1	10
4–7 years	1	10
8–11 years	1	10
12–15 years	3	30
16 years or more	4	40
<b>Which best describes the location of your P–12 school?</b>		
Urban	1	10
Suburban	4	40
Rural	3	30
Not currently working at the P–12 level	2	20
<b>If you are college faculty, are you currently involved in the training/preparation of Japanese language teacher candidates?</b>		
Yes	1	10
No	1	10
Not college faculty	8	80

**STANDARD-SETTING JUDGMENTS**

Table 3 summarizes the standard-setting judgments of panelists (Round 2). 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.<sup>7</sup> 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.

<sup>7</sup> 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).

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 average score is the panel's recommended passing score.

**Table 3**  
*Passing Score Summary by Round of Judgments*

Panelist	Round 1	Round 2
1	62.80	63.00
2	62.60	64.40
3	57.55	55.20
4	53.70	53.30
5	64.40	56.75
6	56.40	55.40
7	60.30	53.00
8	61.60	59.95
9	69.50	67.75
10	54.60	55.10
<b>Average</b>	60.35	58.39
<b>Lowest</b>	53.70	53.00
<b>Highest</b>	69.50	67.75
<b>SD</b>	4.87	5.11
<b>SEJ</b>	1.54	1.62

The panel's passing score recommendation for the *Praxis Japanese: World Language* test is 58.39 (out of a possible 98 raw-score points). The value was rounded to the next highest whole number, 59, to determine the functional recommended passing score. The scale score associated with 59 raw points is 156.

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<sup>8</sup>**

Recommended passing score (CSEM)	Scale score equivalent
59 (4.87)	156
-2 CSEM	50
-1 CSEM	55
+ 1 CSEM	64
+ 2 CSEM	69

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

## 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 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. Nine of the 10 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; 8 of the 10 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 (seven of the 10 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 they recommended; eight of the 10 panelists were *very comfortable*. Nine of the 10 panelists indicated the recommended passing score was *about right* with the remaining panelist indicated that the passing score was *too low*.

<sup>8</sup> 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.

---

## SUMMARY

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis Japanese: World Language* 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 Japanese: World Language* test, the recommended passing score is 59 out of a possible 98 raw-score points. The scale score associated with a raw score of 59 is 156 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.

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: Japanese: World Language (5661)*. Princeton, NJ: Author.

Geisinger, K. F. & McCormick, C. M. (2010). Adopting Cut Scores: Post-Standard-Setting Panel Considerations for Decision Makers. *Educational Measurement: Issues and Practice, 29*: 38–44.

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

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

Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practice, 27*, 15–29.

Tannenbaum, R. J., & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol. 3. Testing and assessment in school psychology and education* (pp. 455–477). Washington, DC: American Psychological Association.

---

## APPENDIX A

### PANELISTS' NAMES & AFFILIATIONS

### ***Participating Panelists With Affiliation***

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Kumiko Gale	Bayside Middle School (VA)
John Hammond	Connections Education dba Pearson Online and Blended Learning (MD)
David House	University of Central Arkansas (AR)
Shunichiro Kurumado	Sunman-Dearborn Community Schools (IN)
Koji Otani	Thomas Jefferson high school for science and technology (VA)
Betsy Paskvan	Dimond High School ASD (AK)
Amanda Ramirez	President - Indiana Foreign Language Teachers Association, Instructor - Avon High School (IN)
William Tai	University of Idaho (ID)
Michael Tsugawa	Michigan City High School (Association of Indiana Teachers of Japanese) (IN)

---

\*One panelist did not wish to be listed in the final report.

---

**APPENDIX B**

**STUDY AGENDA**

## AGENDA

### ***Praxis® Japanese: World Language (5661)* Standard-Setting Study**

#### Day 1

Welcome and Introduction

Overview of Standard Setting and the *Praxis Japanese: World Language Test*

Demonstration of the IME Keyboard

Review the *Praxis Japanese: World Language Test*

Discuss the *Praxis Japanese: World Language Test*

Lunch

Just Qualified Candidate Overview

Define the Knowledge/Skills of a Just Qualified Candidate –  
Small Group Drafts

End of Day 1

## AGENDA

### ***Praxis® Japanese: World Language (5661)* Standard-Setting Study**

Day 2

Overview of Day 2

Define the Knowledge/Skills of a Just Qualified Candidate –  
Whole Group Consensus

Standard-setting Training for Selected-Response Items

Practice Round: Selected-Response Judgments

Lunch

Round 1: Selected-Response Standard-Setting Judgments

End of Day 2

# AGENDA

## ***Praxis® Japanese: World Language (5661)* Standard-Setting Study**

Day 3

Overview of Day 3

Standard-Setting Training for Constructed-Response Items

Practice Round: Constructed-Response Judgments

Round 1: Constructed-Response Standard-Setting Judgments

Lunch

Round 1 Feedback and Round 2 Judgments

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<sup>9</sup>

### A just qualified candidate ...

#### Listening, Reading, and Cultural Knowledge

1. Uses reading strategies such as kanji analysis when encountering moderate to high-frequency vocabulary, inference, and context clues with semi-authentic texts, with kanji up to the third grade level.
2. Has total mastery of basic kana
3. Comprehends a diverse Japanese vocabulary including some commonly used proverbs (kotowaza) and widely used idiomatic expressions, but not slang or regional dialects.
4. Comprehends (a) main ideas and (b) many details in semi-authentic, aural and written communication of varying lengths.
5. Recognizes various registers including formal/informal voice (Honorific / polite / casual) in semi-authentic aural and written communication
6. In listening and reading, is able to distinguish feminine expressions / masculine expressions.
7. Has an understanding of intermediate<sup>10</sup> grammar, syntactical relationships, major forms, and descriptive clauses
8. Has a strong knowledge of Japanese pronunciation including glottal stops, digraphs, and differences between long and short vowels
9. Can identify current, historical, and cultural practices (manners and customs), products (people, places, events, art, music, etc.), and perspectives (values and beliefs).

#### Writing and Speaking

10. Is comprehensible to a native speaker in articulation and pronunciation
11. Is able to model and teach elements of pronunciation including glottal stops, digraphs, and differences between long and short vowels
12. Can express themselves concerning a variety of practical topics as well as express and support opinions on everyday topics, current events, and social issues in contextually appropriate ways
13. Is easily comprehensible to a native speaker, through the use of diverse vocabulary, varied forms, and circumlocution as necessary in writing and speaking
14. Demonstrates basic command and control of mechanics and conventions in writing with respect to orthography, grammar, syntactical relationships, and major forms.
15. Writes and speaks appropriately for various purposes and audiences
16. Organizes ideas to achieve cohesion in writing and speaking

---

<sup>9</sup> Description of the just qualified candidate focuses on the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate.

<sup>10</sup> “Intermediate,” as defined by ACTFL standards.

---

## APPENDIX D

## FINAL EVALUATION RESULTS

**Table D1**  
***Final Evaluation***

	Strongly agree		Agree		Disagree		Strongly disagree	
	N	%	N	%	N	%	N	%
• 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.	9	90	1	10	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	90	1	10	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 between rounds was helpful.	9	90	1	10	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	7	70	2	20	1	10	0	0

**Table D1 (continued)*****Final Evaluation***

<b>How influential was each of the following factors in guiding your standard-setting judgments?</b>	<b>Very influential</b>		<b>Somewhat influential</b>		<b>Not influential</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• The description of the just qualified candidate	8	80	2	20	0	0
• The between-round discussions	9	90	1	10	0	0
• The knowledge/skills required to answer each test item	6	60	4	40	0	0
• The passing scores of other panel members	6	60	4	40	0	0
• My own professional experience	7	70	3	30	0	0
	<b>Very comfortable</b>		<b>Somewhat comfortable</b>		<b>Somewhat uncomfortable</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, how comfortable are you with the panel's recommended passing score?	8	80	2	20	0	0
	0	0	0	0	0	0
	<b>Too low</b>		<b>About right</b>		<b>Too high</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, the recommended passing score is:	1	10	9	90	0	0

## **Background: Overview of the Indiana Score-Setting Workshops**

In 2018 the Indiana Department of Education (IDOE) issued a Request for Proposal (RFP) seeking proposals for a new licensure test program to be effective in 2021. In 2019 the Indiana General Assembly specified that the Indiana State Board of Education would adopt a test program available nationally. In July 2019, the board adopted the *Praxis Series*® from Educational Testing Service (ETS) to be the future licensure test program. IDOE has been working with ETS to be ready to implement current and newly developed tests from its *Praxis Series* to support Indiana's licensure test requirements. These tests will be available in September 2021 to candidates preparing to be Indiana teachers. ETS implemented a process for use in establishing cut scores for newly developed *Praxis* tests and for reviewing multi-state cut scores for current tests for use in Indiana. Each score-setting workshop captured judgements of educators who have expertise in teaching within the targeted content area and experience with new teachers or teacher candidates in the content area. Panelists were approved by the IDOE prior to being invited to serve on the score-setting panel.

Distance-based score-setting workshops were held for 42 *Praxis* tests in 2020. This memorandum reports on one test not included in the 2020 score setting workshops: *Praxis* (0634) American Sign Language Proficiency Interview (ASLPI).

### **Score-Setting for Praxis (0634) ASLPI**

What follows is:

- I. A description of the score-setting process that was implemented
- II. Information about the test
- III. Results of score-setting judgments provided by Indiana educators—including a recommended passing score—along with information about passing scores adopted by other states and by the American Sign Language Teachers Association (ASLTA)

The recommended passing score was 170 on the *Praxis* scale (3+ on the ASLPI score scale). That recommendation—based on judgments from two Indiana ASL educators—agrees with the most widely adopted passing score in other jurisdictions and with the ASLPI score required by ASLTA to earn certification.

#### **I. The Score-Setting Process**

In order to facilitate participation in a field with a small number of practicing teachers, judgments regarding a passing score on ASLPI were captured with an interactive online survey that allowed ASL educators to view videos at different score levels.

The participants were asked to

1. Familiarize themselves with **the proficiency standards for teachers of ASL**, specifically the two language proficiency standards evaluated by the ASLPI: Standards 3 and 4 in Indiana's World Languages Educator Standards addressing Interpretive Communication Skills and Presentational and Interpersonal Communication Skills
2. Familiarize themselves with **the American Sign Language Proficiency Interview** (See "About the ASLPI" below)
3. Familiarize themselves with **ASLPI score scale and Performance Levels** by
  - A. reviewing the Performance Level Descriptors
  - B. reviewing sample video excerpts from performances at the different Performance Levels
4. Provide judgments by bringing together the elements they reviewed in steps 1, 2 and 3 and their own experience and expertise relative to teaching ASL on
  - A. How well aligned ASLPI is as a measure of the skills described in the standards

B. A recommended passing score on the ASLPI for beginning teachers of ASL in Indiana

## II. About the ASLPI

ASLPI is a holistic language evaluation used to determine global ASL proficiency. The evaluation is done via

1. Interview: A video-recorded 20-25 minute face-to-face interactive dialogue between and examinee and an evaluator, and
2. Scoring: A team of evaluators rate the overall proficiency the examinee showed in the interview, scoring on a 0-5 scale through individual rating followed by consensus rating.

The interview is structured to bring out what an individual can do with ASL at the time it is conducted.

Information for examinees about the nature and structure of the interview is provided at [ASLPI preparation](#).

Each recorded interview is rated on the basis of specific criteria in five different categories (which examinees can review here [ASLPI criteria evaluated](#)):

- **Grammar:** linguistic elements including elements conveyed by synchronization of non-manual components (e.g., eye-gaze).
- **Vocabulary:** range and accuracy of vocabulary selection across a variety of topics.
- **Accent/Production:** accuracy of sign formation, i.e., handshape, palm orientation, location and movement as well as clarity of fingerspelled words and numbers.
- **Fluency:** flow of thought that follows a rhythm and pace of delivery that fits the topic.
- **Comprehension:** overall understanding of the conversation. Also examined are the examinee's spontaneity, responses to questions, and ability to provide visual feedback to the interviewer.

ASLPI is scored on a 0-5 scale. Proficiency level descriptors are defined for each of the scores 5, 4, 3, 2, 1, 0, with "in-between" scores available, such as a score of 3+. assigned to a performance that is between the levels 3 and 4. Examinees can review proficiency level descriptors here: [ASLPI Proficiency Levels](#). The levels generally considered for certification are Level 3, Level 4 and Level 3+ shown in this table:

ASLPI level (Praxis score)	Proficiency Level Descriptor
<b>Level 4 (180)</b>	Signers at this proficiency level are able to demonstrate spontaneous elaboration on all familiar and most unfamiliar topics, however, there is incorporation of language patterns other than those of the target language. They are able to use an array of rhetoric (narration, description, argument, and hypothesis) with complex topics in paragraph-length discourse related to employment, current events, and matters of public and community interest. Although they command a good number of grammatical features, they are deficient in some areas such as cohesion, non-manual signals (NMS), and depiction. They are able to present information with sufficient accuracy, clarity, and vocabulary selection to convey intended meaning without misrepresentation or confusion. Comprehension is very good with demonstration of confidence in the discussion of most complex topics.
<b>Level 3+ (170)</b>	Proficiency level between 3 and 4

<b>ASLPI level (Praxis score)</b>	<b>Proficiency Level Descriptor</b>
<b>Level 3 (160)</b>	Signers at this proficiency level are able to express language with sufficient structural accuracy and vocabulary to participate in most familiar and unfamiliar topics about practical, social, and professional situations. They can discuss particular interests with reasonable ease. They demonstrate confidence discussing topics at the paragraph discourse level, but exhibit errors and breakdown when in-depth elaboration and detail is requested. Occasional groping for vocabulary can be present. There is good control of grammar but there are some noticeable imperfections and errors which may interfere with understanding. They tend to function reactively by responding to direct questions or requests for information. They are capable of asking a variety of questions when needed to gather information pertaining to certain situations. They may combine and recombine known language elements to create short paragraph length responses. Their language contains pauses and self-corrections as they search for adequate vocabulary and language forms. Comprehension is often accurate with highly familiar and predictable topics although misunderstandings may occur.

### III. Score setting results

For *Praxis (0634) American Sign Language Proficiency Interview (ASLPI)*, the recommended study value (RSV) is 170, which is the median adopted score (MAS) for the test among the 12 states using it and equivalent to 3+ on the ASLPI score scale.

In the format such results were reported for other adopted tests, the details of score setting for ASLPI are as follows:

**Table 1. Indiana Score-Setting Schedule with Link to Preparation Material**

Test Code	Praxis Test Title	Score-setting Date
0634	<a href="#">American Sign Language Proficiency Interview</a>	Mar 2021

**Table 2. A Praxis Title where the IN Score-Setting Panel Recommended the Median Adopted Score**

Test Name (Code)	Panel Size	RSV=MAS
American Sign Language Proficiency Interview (0634)	2	160 170 180
	3	3+ 4

Two ASL teachers participated in the study (out of four who initially responded to the e-mail invitation initially sent to 15).<sup>1</sup> Besides the elements they reviewed as described in “The Score Setting Process” above, the panelists were informed that:

1. Among the states currently using ASLPI through Praxis for licensure, different states have set different passing scores:  
Level 3—160 on the Praxis scale—is the minimum passing score for two states (ID and WY)  
Level 3+—170—is the passing score for nine states (AR, CT, HI, KY, LA, PA, RI, TN, and VT)  
Level 4—180—is the passing score for one state (WV).

and that:

<sup>1</sup> As proportions of the targeted population these response rates are quite good to e-mail invitations, but, with an aim for a better response, three rounds of reminders were sent.

2. The American Sign Language Teachers Association has two levels of certification and has ASLPI passing scores at Level 3+ for the CERTIFIED level and at Level 4 for the MASTER level (ASLTA certification)

Before being asked for a recommended score, the panelists were asked

- Taking into account Standards 3 and 4 for Interpretive Communication Skills and Presentational and Interpersonal Communication Skills you reviewed in STEP 1, and based on your review of ASLPI, how would you judge ASLPI's alignment as a measure of those skills?

Both indicated alignment: One "Very well aligned," and one "Well aligned," with the other options being "Moderately aligned" and "Not aligned."

The panelists were then asked

- What passing score would you recommend on the ASLPI for beginning teachers of ASL in Indiana (In terms of ASL skills what ASLPI Proficiency Level would the minimally qualified beginning teacher be at)? Available responses were Level 5 (200 on the Praxis scale),, Level 4+ (190), Level 4 (180), Level 3+ (170), Level 3 (160), Level 2+ (150), and Level 2 (140).

One selected Level 3 (160) and one Level 4 (180) with the median at Level 3+ (170). The panelists provided the following explanatory comments:

- "For beginning teachers, a level 3 is sufficient enough to teach at the k-12 level. The more years a teacher is teaching the language, the more their skills will improve so a level 3+ or 4 for ASLTA certification is a goal the teacher can set for themselves but beginning teachers should be at a level 3."
- "Standard 3 and Standard 4, states the ASL teachers would demonstrate their skills at the level equivalent to the intermediate-high level as defined by ACTFL's proficiency guidelines. These standard requirements would represent ASLPI proficiency level of 5, 4+, and 4. If Indiana were to adopt, ASLTA's certification with ASLP passing scores of Level 3+, this would be appropriate for teachers of ASL teaching beginning level courses, not immediate to advance ASL courses."



## Multistate Standard-Setting Technical Report

### **PRAXIS® MUSIC: VOCAL AND GENERAL KNOWLEDGE (5116)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

April 2021

---

# EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Music: Vocal and General Knowledge (5116) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

## PARTICIPATING STATES

Panelists from seven states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music 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* Music: Vocal and General Knowledge test, the recommended passing score is 62 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 62 is 153 on a 100–200 scale.

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*<sup>®</sup> Music: Vocal and General Knowledge (5116) test, research staff from ETS designed and conducted a multistate standard-setting study in March 2021. Education agencies<sup>1</sup> recommended panelists with (a) experience as either music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music teachers. Seven states (Table 1) were represented by 12 panelists. (See Appendix A for the names and affiliations of the panelists.)

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

Arkansas (1 panelist)	Nevada (1 panelist)
Indiana (1 panelist)	Virginia (2 panelists)
Kansas (2 panelists)	West Virginia (2 panelists)
North Carolina (3 panelists)	

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

ETS provides a recommended passing score from the multistate standard-setting study to education agencies. In each state, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in accordance with applicable regulations. This study provides a recommended passing score, which represents the combined judgments of a group of experienced educators. Each state may want to consider the recommended passing score but also other sources of information when setting the final *Praxis* Music: Vocal and General 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 error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the *Praxis* Music: Vocal and General Knowledge test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a state to recognize that any test score on any

<sup>1</sup> States and jurisdictions that currently use *Praxis* tests were invited to participate in the multistate standard-setting study.

standardized test—including a *Praxis* Music: Vocal and General Knowledge test score—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows a state 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 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 possesses the required knowledge/skills. The state needs to consider which decision error is more important to minimize.

---

# OVERVIEW OF THE *PRAXIS*<sup>®</sup> MUSIC: VOCAL AND GENERAL KNOWLEDGE TEST

The *Praxis*<sup>®</sup> Music: Vocal and General Knowledge *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test is designed to assess a beginning music teacher’s knowledge of music and music education.

The two-hour and five minute assessment contains 125 selected-response items<sup>2</sup> covering five content areas: *Music History and Literature* (approximately 16 items), *Music Theory and Composition* (approximately 18 items), *Music Performance* (approximately 25 items), *Music Pedagogy and Instructional Practices* (approximately 49 items), and *Professional Issues and Music Technology* (approximately 17 items).<sup>3</sup> The reporting scale for the *Praxis* Music: Vocal and General Knowledge 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 agenda for the panel meeting.

### REVIEWING THE TEST

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

---

<sup>2</sup> Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate’s score.

<sup>3</sup> 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.

The panel created a description of the just qualified candidate —the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, the panel first split into smaller groups to consider the just qualified candidate. The full panel then 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 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 Music: Vocal and General Knowledge* test was a probability-based Modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). In this study, each panelist judged each item on the likelihood (probability or chance) that the 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.

---

## RESULTS

### EXPERT PANELS

Table 2 presents a summary of the panelists' demographic information. The panel included 12 educators representing seven states. (See Appendix A for a listing of panelists.) Ten panelists were teachers and two were college faculty. All of the faculty members' job responsibilities included the training of music teachers.

**Table 2**  
*Panel Member Demographics*

	<i>N</i>	<i>%</i>
<b>Current position</b>		
Teacher	10	73
College faculty	2	17
<b>Race</b>		
Black or African American	1	8
Hispanic or Latino	1	8
White	10	83
<b>Gender</b>		
Female	6	50
Male	6	50
<b>Are you currently certified to teach music in your state?</b>		
Yes	11	92
No	1	8
<b>Are you currently teaching music in your state?</b>		
Yes	12	100
No	0	0
<b>Are you currently supervising or mentoring other music teachers?</b>		
Yes	7	58
No	5	42
<b>At what K–12 grade level are you currently teaching this subject?</b>		
Elementary (K–5 or K–6)	4	33
Middle school (6–8 or 7–9)	2	17
Middle and High school	1	8
High school (9–12 or 10–12)	3	25
Not currently teaching at the P–12 level	2	17

**Table 2 (continued)****Panel Member Demographics**

	<i>N</i>	<i>%</i>
<b>Including this year, how many years of experience do you have teaching music?</b>		
3 years or less	2	17
4–7 years	2	17
8–11 years	0	0
12–15 years	0	0
16 years or more	8	67
<b>Which best describes the location of your K–12 school?</b>		
Urban	0	0
Suburban	5	42
Rural	5	42
Not currently working at the K–12 level	2	17
<b>If you are college faculty, are you currently involved in the training/preparation of teacher candidates in this subject?</b>		
Yes	3	25
No	0	0
Not college faculty	9	75

**STANDARD-SETTING JUDGMENTS**

Table 3 summarizes the standard-setting judgments of panelists. The table shows the passing scores—the number of raw points needed to pass the test—recommended by each panelist. Table 4 also presents the summary data but with the highest outlier removed. Although two additional outliers were detected during our routine outlier analysis (High, 2000), ETS does not recommend that their data be removed because they were believed to be following the standard-setting process faithfully. The panelist who provided the highest scores, was believed to be making judgments that would raise the mean, instead of making judgments based on the standard-setting training.

Tables 3 and 4 include the estimate 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.<sup>4</sup> 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.

<sup>4</sup> 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).

Round 1 judgments are made without discussion among the panelists. The most variability in judgments, therefore, is typically present in the first round. Round 2 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease — indicating convergence among the panelists' judgments — was observed (see Tables 3 and 4). ETS recommends that the Round 2 mean recommendation in Table 4 be considered as the panel's recommended passing score.

**Table 3**  
*Passing Score Summary by Round of Judgments*

Panelist	Round 1	Round 2
1	58.80	59.20
2	61.35	61.75
3	69.60	66.20
4	60.95	61.05
5	64.00	64.05
6	88.95	81.50
7	77.50	74.85
8	59.00	56.60
9	62.35	63.45
10	61.60	61.05
11	65.60	62.40
12	44.90	46.60
<b>Average</b>	64.55	63.23
<b>Lowest</b>	44.90	46.60
<b>Highest</b>	88.95	81.50
<b>SD</b>	10.77	8.67
<b>SEJ</b>	3.11	2.50

**Table 4**  
*Passing Score Summary with Highest Outlier Removed*

Statistic	Round 1	Round 2
<b>Average</b>	62.33	61.56
<b>Lowest</b>	44.90	46.60
<b>Highest</b>	77.50	74.85
<b>SD</b>	7.92	6.81
<b>SEJ</b>	2.39	2.05

With the removal of the highest outlier, the panel's passing score recommendation for the *Praxis Music: Vocal and General Knowledge* test is 61.56 (out of a possible 100 raw-score points). The value was rounded to the next highest whole number, 62, to determine the functional recommended passing score. The scale score associated with 62 raw points is 153.

Table 5 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. (With all of the data, the scale score associated with 64 raw points is 156, with a CSEM of 4.82.)

**Table 5**

***Passing Scores Within 1 and 2 CSEM of the Recommended Passing Score<sup>5,6</sup>***

<b>Recommended passing score (CSEM)</b>	<b>Scale score equivalent</b>
62 (4.88)	153
-2 CSEM	53
-1 CSEM	58
+ 1 CSEM	67
+ 2 CSEM	72

**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

<sup>5</sup> 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.

<sup>6</sup> The recommended passing score, and associated data, shown in this table does not include the highest outlier.

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 between-round discussions were at least *somewhat influential* in guiding their judgments. More than half of the panelists (10 of the 12 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

The panel was presented with the passing score that was calculated during the study, 64 out of 100 raw score points. The recommendation to remove the outlier occurred after the meeting. Based on the score of 64 out of 100, all of the panelists indicated they were at least *somewhat comfortable* with the recommended passing score; nine of the 14 panelists were *very comfortable*. Ten of the 12 panelists indicated the recommended passing score was *about right*; one panelist indicated that the passing score was *too low* and one panelist indicated 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* Music: Vocal and General Knowledge test, research staff from ETS designed and conducted a multistate standard-setting study.

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the *Praxis* Music: Vocal and General Knowledge test, the recommended passing score is 62 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 62 is 153 on a 100–200 scale.

---

## REFERENCES

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

ETS. (in press). *The Praxis Series®: The Praxis Study Companion: Music: Vocal and General Knowledge (5116)*. 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>

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***

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Eric Baskerville	Chesapeake Public Schools (VA)
Cheryl Bates	Henderson State University (AR)
Kim Cantley	Marsh Fork Elementary School (WV)
Darrell Crowther	Depoali Middle School (NV)
Ben Kambs	Fairfield High School (IN)
Sarah Kerr	Elkin Elementary School (NC)
Joey Nickols	Colby High School (KS)
Christopher Pringle	Charlotte Mecklenburg Schools (NC)
Ellen Ramey	George Washington High School (WV)
Dana Van Slyke	Virginia Choral Director's Assoc./ FCPS (VA)
DeAnna Walker	Surry County Schools (NC)
Tom Wine	Wichita State University (KS)

---

---

**APPENDIX B**

**STUDY AGENDA**

## AGENDA

### ***Praxis® Music: Vocal and General Knowledge (5116)* Standard-Setting Study**

Day 1

Welcome and Introduction

Overview of Standard Setting and the *Praxis Music: Vocal and General Knowledge Test*

Review the *Praxis Music: Vocal and General Knowledge Test*

Discuss the *Praxis Music: Vocal and General Knowledge Test*

Lunch

Just Qualified Candidate Overview

Break

Define the Knowledge/Skills of a Just Qualified Candidate

End of Day 1

## **AGENDA**

### ***Praxis® Music: Vocal and General Knowledge (5116)* Standard-Setting Study**

Day 2

Overview of Day 2

Define the Knowledge/Skills of a Just Qualified Candidate  
(continued)

Break

Standard-setting training presentation

Practice Round: Selected-response standard-setting judgments

Break

Practice Round: Data Discussion

Lunch

Round 1 Feedback and Round 2 Judgments

End of Day 2

## AGENDA

### ***Praxis® Music: Vocal and General Knowledge (5116)* 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)

Lunch

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<sup>7</sup>

### A just qualified candidate...

#### ***Music History & Theory***

1. Understands periods of Western music history and genres (including popular music genres)
2. Knows different genres and composers by aural identification.
3. Understands Western music theory, forms, and compositional practices
4. Is familiar with world musics, instruments, and timbres by regions and their influences on other genres

#### ***Performance***

5. Knows instructional techniques for vocal/instrumental rehearsal, warm-up, tuning, and performance.
6. Is familiar with identifying and managing proper balance and blend, choral diction, dynamics and intonation
7. Is familiar with effective strategies for programming and presentation of performances
8. Understands basic accompaniment techniques, including the selection of appropriate accompaniment.
9. Knows how to adapt an accompaniment for varied purposes
10. Knows basic beat patterns, conducting gestures, and cueing.

#### ***Instruction***

11. Knows strategies for instruction, classroom management and assessments for diverse learners and contexts
12. Knows local, state, and national standards
13. Knows basic techniques for improvisation, composition and arranging and how to teach them through performance literature
14. Knows common approaches for fostering musical expression and creativity.

#### ***Professional Responsibilities and Technology***

15. Is familiar with professional and administrative responsibilities, and situational factors affecting music instruction and advocacy
16. Is familiar with copyright laws and legal issues specific to teaching music.
17. Is familiar with professional organizations
18. Knows how to utilize basic technology for performance, instruction, recording
19. Understands ethical and safety issues, including personal social media and technology usage

---

<sup>7</sup> 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...

#### *Instructional Activities*

20. Knows how to select developmentally relevant repertoire
21. Knows common pedagogical and instructional techniques and how to implement them
22. Knows the logistics of concert planning
23. Knows strategies for developing sight-reading skills and music literacy

---

## APPENDIX D

## FINAL EVALUATION RESULTS

**Table D1**  
***Final Evaluation***

	Strongly agree		Agree		Disagree		Strongly disagree	
	N	%	N	%	N	%	N	%
• I understood the purpose of this study.	11	92	1	8	0	0	0	0
• The instructions and explanations provided by the facilitators were clear.	11	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.	12	100	0	0	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	11	92	1	8	0	0	0	0
• The opportunity for feedback and discussion between rounds was helpful.	11	92	1	8	0	0	0	0
• The process of making the standard-setting judgments was easy to follow.	9	75	3	25	0	0	0	0

**Table D1 (continued)*****Final Evaluation***

<b>How influential was each of the following factors in guiding your standard-setting judgments?</b>	<b>Very influential</b>		<b>Somewhat influential</b>		<b>Not influential</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• The description of the just qualified candidate	12	100	0	0	0	0
• The between-round discussions	9	75	3	25	0	0
• The knowledge/skills required to answer each test item	9	75	3	25	0	0
• The passing scores of other panel members	2	17	10	83	0	0
• My own professional experience	10	83	2	17	0	0
	<b>Very comfortable</b>		<b>Somewhat comfortable</b>		<b>Somewhat uncomfortable</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, how comfortable are you with the panel's recommended passing score?	9	75	3	25	0	0
	0	0	0	0	0	0
	<b>Too low</b>		<b>About right</b>		<b>Too high</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, the recommended passing score is:	1	8	10	83	1	8



## Multistate Standard-Setting Technical Report

### **PRAXIS® MUSIC: INSTRUMENTAL AND GENERAL KNOWLEDGE (5115)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

April 2021

---

# EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the *Praxis*® Music: Instrumental and General Knowledge (5115) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

## PARTICIPATING STATES

Panelists from nine states were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music 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* Music: Instrumental and General Knowledge test, the recommended passing score is 60 out of a possible 100 raw-score points. The scale score associated with a raw score of 60 is 150 on a 100–200 scale.

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Music: Instrumental and General Knowledge (5115) test, research staff from ETS designed and conducted a multistate standard-setting study in March 2021. Education agencies<sup>1</sup> recommended panelists with (a) experience as either music teachers or college faculty who prepare music teachers and (b) familiarity with the knowledge and skills required of beginning music teachers. Nine states (Table 1) were represented by 13 panelists. (See Appendix A for the names and affiliations of the panelists.)

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

Arkansas (2 panelists)	Nevada (1 panelist)
Indiana (2 panelists)	Utah (1 panelist)
Kansas (2 panelists)	Virginia (2 panelists)
Kentucky (1 panelist)	West Virginia (1 panelist)
North Carolina (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* Music: Instrumental and General 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 error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the *Praxis* Music: Instrumental and General Knowledge test score and the latter, the reliability of

<sup>1</sup> States and jurisdictions that currently use *Praxis* tests were invited to participate in the multistate standard-setting study.

panelists' passing-score recommendation. The SEM allows a state to recognize that any test score on any standardized test—including a *Praxis* Music: Instrumental and General Knowledge test score—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows a state 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 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 possesses the required knowledge/skills. The state needs to consider which decision error is more important to minimize.

---

# OVERVIEW OF THE *PRAXIS*<sup>®</sup> MUSIC: INSTRUMENTAL AND GENERAL KNOWLEDGE TEST

The *Praxis*<sup>®</sup> Music: Instrumental and General Knowledge Test *at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test is designed to assess a beginning music teacher’s knowledge of music and music education.

The two-hour and five minute assessment contains 125 selected-response items<sup>2</sup> covering five content areas: *Music History and Literature* (approximately 16 items), *Music Theory and Composition* (approximately 18 items), *Music Performance* (approximately 25 items), *Music Pedagogy and Instructional Practices* (approximately 49 items), and *Professional Issues and Music Technology* (approximately 17 items).<sup>3</sup> The reporting scale for the *Praxis* Music: Instrumental and General Knowledge 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 agenda for the panel meeting.

### REVIEWING THE TEST

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

---

<sup>2</sup> Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate’s score.

<sup>3</sup> 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.

The panel created a description of the just qualified candidate —the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, the panel first split into smaller groups to consider the just qualified candidate. The full panel then 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 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 Music: Instrumental and General Knowledge* test was a probability-based Modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). In this study, each panelist judged each item on the likelihood (probability or chance) that the 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.

---

# RESULTS

## EXPERT PANELS

Table 2 presents a summary of the panelists' demographic information. The panel included 13 educators representing nine states. (See Appendix A for a listing of panelists.) Nine panelists were teachers, two were college faculty, one was an administrator or department head, and one was a state educational specialist. Two of the three faculty members' job responsibilities included the training of music teachers.

**Table 2**  
*Panel Member Demographics*

	<i>N</i>	<i>%</i>
<b>Current position</b>		
Teacher	9	69
Administrator/Department head	1	8
College faculty	2	15
State Educational Specialist	1	8
<b>Race</b>		
Black or African American	2	15
White	10	77
Native American, German, Scottish, & Irish	1	8
<b>Gender</b>		
Female	7	54
Male	6	46
<b>Are you currently certified to teach music in your state?</b>		
Yes	11	85
No	2	15
<b>Are you currently teaching music in your state?</b>		
Yes	11	85
No	2	15
<b>Are you currently supervising or mentoring other music teachers?</b>		
Yes	8	62
No	5	38

**Table 2 (continued)****Panel Member Demographics**

	<i>N</i>	<i>%</i>
<b>At what K–12 grade level are you currently teaching music?</b>		
Elementary (K–5 or K–6)	3	23
Middle school (6–8 or 7–9)	2	15
Middle and High school	1	8
High school (9–12 or 10–12)	2	15
All Grades	1	8
Other	1	8
Not currently teaching at the K–12 level	3	23
<b>Including this year, how many years of experience do you have teaching music?</b>		
3 years or less	0	0
4–7 years	1	8
8–11 years	2	15
12–15 years	5	38
16 years or more	5	38
<b>Which best describes the location of your K–12 school?</b>		
Urban	6	46
Suburban	4	31
Rural	3	23
Not currently working at the K–12 level	0	0
<b>If you are college faculty, are you currently involved in the training/preparation of music teacher candidates?</b>		
Yes	2	15
No	0	0
Not college faculty	11	85

**STANDARD-SETTING JUDGMENTS**

Table 3 summarizes the standard-setting judgments of panelists. The table shows the passing scores—the number of raw points needed to pass the test—recommended by each panelist.

Table 3 also includes the estimate 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.<sup>4</sup> It indicates how likely

<sup>4</sup> 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).

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.

Round 1 judgments are made without discussion among the panelists. The most variability in judgments, therefore, is typically present in the first round. Round 2 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease — indicating convergence among the panelists' judgments — was observed (see Table 3). The Round 2 average score is the panel's recommended passing score.

**Table 3**  
*Passing Score Summary by Round of Judgments*

<b>Panelist</b>	<b>Round 1</b>	<b>Round 2</b>
1	55.90	57.00
2	59.40	59.80
3	65.95	65.45
4	49.80	53.80
5	57.60	58.90
6	62.70	63.80
7	52.70	53.00
8	58.40	59.00
9	59.75	60.75
10	51.90	55.50
11	64.25	63.55
12	53.10	54.10
13	69.20	68.40
<b>Average</b>	58.51	59.47
<b>Lowest</b>	49.80	53.00
<b>Highest</b>	69.20	68.40
<b>SD</b>	5.86	4.81
<b>SEJ</b>	1.63	1.33

The panel's passing score recommendation for the *Praxis* Music: Instrumental and General Knowledge test is 59.47 (out of a possible 100 raw-score points). The value was rounded to the next highest whole number, 60, to determine the functional recommended passing score. The scale score associated with 60 raw points is 150.

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<sup>5</sup>***

<b>Recommended passing score (CSEM)</b>	<b>Scale score equivalent</b>
60 (4.92)	150
-2 CSEM	51
-1 CSEM	56
+ 1 CSEM	65
+ 2 CSEM	70

***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; 11 of the 13 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 (11 of the 13 panelists) indicated that their own professional experience was *very influential* in guiding their judgments.

All indicated they were at least *somewhat comfortable* with the passing score they recommended; 11 of the 13 panelists were *very comfortable*. All of the 13 panelists indicated the recommended passing score was *about right*.

---

<sup>5</sup> 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.

---

## SUMMARY

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis* Music: Instrumental and General Knowledge test, research staff from ETS designed and conducted a multistate standard-setting study.

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the *Praxis* Music: Instrumental and General Knowledge test, the recommended passing score is 60 out of a possible 100 raw-score points. The scale score associated with a raw score of 60 is 150 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.

ETS. (in press). *The Praxis Series®: The Praxis Study Companion: Music: Instrumental and General Knowledge (5115)*. 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***

---

<b><u>Panelist</u></b>	<b><u>Affiliation</u></b>
Daniel Abrahams	University of Arkansas - Fayetteville (AR)
Lauren Bulla	Shining Rock Classical Academy (NC)
Eric Colgrove	Harding University (AR)
Patrick Collins	Mojave High School (NV)
William Cosby	Wood County (WV)
Andrea Dinkel	Galena USD 499 (KS)
David Fonda	Camargo Elementary School (KY)
Christopher Moseley	Henrico County Public Schools (VA)
Susan E. Sieler	Shortridge High School (IN)
Amanda Vanausdall	USD 259 Wichita Public Schools (KS)
Susan Walker	Lasalle Academy (IN)
Sherry Wozniak	Isle of Wight County Schools (VA)

---

\*One panelist did not wish to be listed in the final report.

---

**APPENDIX B**

**STUDY AGENDA**

## AGENDA

### ***Praxis® Music: Instrumental and General Knowledge (5115)* Standard-Setting Study**

#### Day 1

Welcome and Introduction

Overview of Standard Setting and the *Praxis Music: Instrumental and General Knowledge Test*

Review the *Praxis Music: Instrumental and General Knowledge Test*

Discuss the *Praxis Music: Instrumental and General Knowledge Test*

Lunch

Just Qualified Candidate Overview

Break

Define the Knowledge/Skills of a Just Qualified Candidate

End of Day 1

## AGENDA

### ***Praxis® Music: Instrumental and General Knowledge (5115)*** **Standard-Setting Study**

Day 2

Overview of Day 2

Define the Knowledge/Skills of a Just Qualified Candidate  
(continued)

Break

Standard-setting training presentation

Practice Round: Selected-response standard-setting judgments

Break

Practice Round: Data Discussion

Lunch

Round 1: Selected-response standard-setting judgments

End of Day 2

## AGENDA

### ***Praxis® Music: Instrumental and General Knowledge (5115)* 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)

Lunch

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<sup>6</sup>**

### **A just qualified candidate...**

#### ***Music History & Literature; Music Theory and Compositions***

1. Knows period of music history and genres
2. Understands the elements of music and theory (e.g. transposition, ear training/aural skills.) theory and compositional practices
3. Understands elements of compositional practices (e.g. scoring/arranging)
4. Is familiar with global musics

#### ***Music Performance***

5. Understands performance techniques for general music and instrumental rehearsal and performance
6. Understands conducting techniques in rehearsal and performance
7. Knows warm-up, tuning, and intonation practices
8. Knows best practices for programming and presentation/facilitation of performances
9. Knows effective use of aural skills to detect errors in an ensemble setting
10. Familiar with basic accompaniment techniques
11. Understands proper rehearsal and concert etiquette (students and audience)

#### ***Music Pedagogy and Instructional Practices***

12. Knows best practices for instruction, assessment, and accommodations for differentiating instruction
13. Knows best practices for classroom management
14. Knows national standards
15. Knows how to teach basic techniques of improvisation, composition and arranging through performance
16. Familiar with approaches for fostering musical expression (performance and creativity)
17. Knows and promotes the care and maintenance of instruments and materials
18. Knows common pedagogical approaches (e.g. Kodaly, Dalcroze, Orff, Gordon)
19. Familiar with vocal techniques
20. Understands how to select appropriate repertoire
21. Knows how to address common instrumental and ensemble performance challenges and their causes
22. Understands strategies for developing music literacy
23. Understands strategies for developing individual and group performance skills

---

<sup>6</sup> 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...**

##### ***Professional Issues and Music Technology***

24. Familiar with professional resources and organizations (national and state organizations)
25. Is familiar with professional ethics, advocacy, copyright and legal issues, and responsibilities specific to teaching music
26. Is familiar with the external influences that affect the music program, curriculum, and student participation
27. Knows appropriate use of technology for performance, instruction, recording
28. Understands ethical and safety issues

---

## APPENDIX D

## FINAL EVALUATION RESULTS

**Table D1**  
***Final Evaluation***

	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.	12	92	1	8	0	0	0	0
• The explanation of how the recommended passing score is computed was clear.	8	62	5	38	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.	9	69	4	31	0	0	0	0

**Table D1 (continued)*****Final Evaluation***

<b>How influential was each of the following factors in guiding your standard-setting judgments?</b>	<b>Very influential</b>		<b>Somewhat influential</b>		<b>Not influential</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• The description of the just qualified candidate	11	85	2	15	0	0
• The between-round discussions	10	77	3	23	0	0
• The knowledge/skills required to answer each test item	10	77	3	23	0	0
• The passing scores of other panel members	2	15	10	77	1	8
• My own professional experience	11	85	2	15	0	0
	<b>Very comfortable</b>		<b>Somewhat comfortable</b>		<b>Somewhat uncomfortable</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, how comfortable are you with the panel's recommended passing score?	11	85	2	15	0	0
	0	0	13	100	0	0
	<b>Too low</b>		<b>About right</b>		<b>Too high</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
• Overall, the recommended passing score is:	0	0	13	100	0	0