

OFFICE OF CHIEF OPERATIONS OFFICER
Summary of State Board of Education Items
Consent Agenda
June 19-20, 2014

OFFICE OF EDUCATOR LICENSURE

- J. Approval of Praxis II Tests and recommended passing scores as recommended by the Commission on Teacher and Administrator Education, Certification and Licensure and Development
(Has cleared the Administrative Procedures Act process with no public comment)
1. Praxis II Elementary Education: Curriculum, Instruction and Assessment (5017) to Replace 0011/5011
 2. Praxis II Elementary Education: Content Knowledge (5018)
 3. Praxis II Middle School Science (5440)

Background Information:

Educational Testing Service (ETS) has concluded several multistate standard-setting studies of Praxis II tests accepted by the Mississippi Department of Education. As a result of the studies, the current tests will be discontinued by ETS and replaced with the newly regenerated tests beginning in September of 2014.

On March 7, 2014, the Commission on Teacher and Administrator Education Certification and Licensure and Development approved the recommendation of ETS to accept the Praxis II tests and recommended passing scores.

All Praxis scores submitted to the State Board of Education for approval reflect passing scores recommended by the multistate standard-setting panel.

Recommendation: Approval

Back-up material attached

1. Praxis II Elementary Education: Curriculum, Instruction and Assessment
(5017) to Replace 0011/5011



Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

**PRAXIS™ ELEMENTARY EDUCATION:
CURRICULUM, INSTRUCTION, AND ASSESSMENT (5017)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2014

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ Elementary Education: Curriculum, Instruction, and Assessment (5017) 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 elementary school teachers or college faculty who prepare elementary school teachers and (b) familiarity with the knowledge and skills required of beginning elementary school 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 Elementary Education: Curriculum, Instruction, and Assessment test, the recommended passing score is 62 out of a possible 100 raw-score points. The scaled 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™ Elementary Education: Curriculum, Instruction, and Assessment (5017) test, research staff from ETS designed and conducted a multistate standard-setting study in January 2014 in Princeton, New Jersey. Education agencies¹ recommended panelists with (a) experience as either elementary school teachers or college faculty who prepare elementary school teachers and (b) familiarity with the knowledge and skills required of beginning elementary school teachers. Seven states (Table 1) were represented by 15 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating States and Number of Panelists

Maryland (3 panelists)	Nevada (1 panelist)
Mississippi (2 panelists)	Tennessee (1 panelist)
North Dakota (2 panelists)	Wyoming (3 panelists)
Nebraska (3 panelists)	

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

ETS provides a recommended passing score from the multistate standard-setting study to education agencies. In each state, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in accordance with applicable regulations. This study provides a recommended passing score, which represents the combined judgments of a group of experienced educators. Each state may want to consider the recommended passing score but also other sources of information when setting the final Praxis Elementary Education: Curriculum, Instruction, and Assessment passing score (see Geisinger & McCormick, 2010). A state may accept the recommended passing score, adjust the score upward to reflect more stringent expectations, or adjust the score downward to reflect more lenient expectations. There is no *correct* decision; the appropriateness of any adjustment may only be evaluated in terms of its meeting the state’s needs.

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

Two sources of information to consider when setting the passing score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the Praxis Elementary Education: Curriculum, Instruction, and Assessment 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 standardized test—including a Praxis Elementary Education: Curriculum, Instruction, and Assessment 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 does possess the required knowledge/skills. The state needs to consider which decision error is more important to minimize.

OVERVIEW OF THE PRAXIS ELEMENTARY EDUCATION: CURRICULUM, INSTRUCTION, AND ASSESSMENT TEST

The Praxis Elementary Education: Curriculum, Instruction, and Assessment *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level elementary school teachers have the knowledge/skills believed necessary for competent professional practice.

The two-hour and 10 minute test contains 120 selected-response items² covering five content areas: *Reading and Language Arts* (approximately 37 items), *Mathematics* (approximately 31 items), *Science* (approximately 20 items), *Social Studies* (approximately 17 items), and *Art, Music, and Physical Education* (approximately 15 items).³ The reporting scale for the Praxis Elementary Education: Curriculum, Instruction, and Assessment test ranges from 100 to 200 scaled-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 reviewed 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.

² Twenty of the 120 selected-response items are pretest items and do not contribute to a candidate's score.

³ The number of items for each content area may vary slightly from form to form of the test.

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 Elementary Education: Curriculum, Instruction, and Assessment 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 decided if, overall, the item would be difficult for the just qualified candidate, easy for the just qualified candidate or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Difficult items for the just qualified candidate are in the 0 to .30 range.
- Moderately difficult/easy items for the just qualified candidate are in the .40 to .60 range.
- Easy items for the just qualified candidate are in the .70 to 1 range.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that an item would be easy for the just qualified candidate, the initial decision located the item in the .70 to 1 range. The second decision for the panelist was to decide if the likelihood of answering it correctly is .70, .80, .90, .95 or 1.

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

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 15 educators representing seven states. (See Appendix A for a listing of panelists.) Twelve panelists were teachers, two were college faculty, and one held another position. The two faculty members' job responsibilities included the training of elementary school teachers.

2. Praxis II Elementary Education: Content Knowledge (5018)



Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

**PRAXIS™ ELEMENTARY EDUCATION:
CONTENT KNOWLEDGE (5018)**

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2014

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ Elementary Education: Content Knowledge (5018) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING STATES

Panelists from seven states and Guam were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either elementary school teachers or college faculty who prepare elementary school teachers and (b) familiarity with the knowledge and skills required of beginning elementary school 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 Elementary Education: Content Knowledge test, the recommended passing score is 83 out of a possible 120 raw-score points. The scaled score associated with a raw score of 83 is 163 on a 100–200 scale.

To support the decision-making process for education agencies establishing a passing score (cut score) for the Praxis™ Elementary Education: Content Knowledge (5018) test, research staff from ETS designed and conducted a multistate standard-setting study in January 2014 in Princeton, New Jersey. Education agencies¹ recommended panelists with (a) experience as either elementary school teachers or college faculty who prepare elementary school teachers and (b) familiarity with the knowledge and skills required of beginning elementary school teachers. Seven states and Guam (Table 1) were represented by 16 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating Jurisdictions and Number of Panelists

Delaware (1 panelist)	Mississippi (3 panelists)
Guam (1 panelist)	Montana (2 panelists)
Iowa (2 panelists)	Nevada (2 panelists)
Louisiana (2 panelists)	South Dakota (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 jurisdiction, 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 jurisdiction may want to consider the recommended passing score but also other sources of information when setting the final Praxis Elementary Education: Content Knowledge passing score (see Geisinger & McCormick, 2010). A jurisdiction 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 jurisdiction’s needs.

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

Two sources of information to consider when setting the passing score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the Praxis Elementary Education: Content Knowledge test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a jurisdiction to recognize that any test score on any standardized test—including a Praxis Elementary Education: Content 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 jurisdiction 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 jurisdiction should consider the likelihood of classification errors. That is, when adjusting a passing score, policymakers should consider whether it is more important to minimize a false-positive decision or to minimize a false-negative decision. A false-positive decision occurs when a candidate's test score suggests that he should receive a license/certificate, but his actual level of knowledge/skills indicates otherwise (i.e., the candidate does not possess the required knowledge/skills). A false-negative decision occurs when a candidate's test score suggests that she should not receive a license/certificate, but she actually does possess the required knowledge/skills. The jurisdiction needs to consider which decision error is more important to minimize.

OVERVIEW OF THE PRAXIS ELEMENTARY EDUCATION: CONTENT KNOWLEDGE TEST

The Praxis Elementary Education: Content Knowledge *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level elementary school teachers have the knowledge/skills believed necessary for competent professional practice.

The two and a half-hour test contains 140 selected-response items² covering four content areas: *Reading and Language Arts* (approximately 49 items), *Mathematics* (approximately 41 items), *Social Studies* (approximately 25 items), and *Science* (approximately 25 items).³ The reporting scale for the Praxis Elementary Education: Content Knowledge test ranges from 100 to 200 scaled-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 reviewed 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.

² Twenty of the 140 selected-response and numeric entry items are pretest items and do not contribute to a candidate's score.

³ The number of items for each content area may vary slightly from form to form of the test.

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 Elementary Education: Content 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 decided if, overall, the item would be difficult for the just qualified candidate, easy for the just qualified candidate or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Difficult items for the just qualified candidate are in the 0 to .30 range.
- Moderately difficult/easy items for the just qualified candidate are in the .40 to .60 range.
- Easy items for the just qualified candidate are in the .70 to 1 range.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that an item would be easy for the just qualified candidate, the initial decision located the item in the .70 to 1 range. The second decision for the panelist was to decide if the likelihood of answering it correctly is .70, .80, .90, .95 or 1.

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

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 16 educators representing seven states and Guam. (See Appendix A for a listing of panelists.) Seven panelists were teachers, six were college faculty, two were administrators or department heads, and one held another position. All six faculty members' job responsibilities included the training of elementary school teachers.

3. Praxis II Middle School Science (5440)



Listening. Learning. Leading.

Multistate Standard-Setting Technical Report

PRAXIS™ MIDDLE SCHOOL SCIENCE (5440)

Licensure and Credentialing Research

ETS

Princeton, New Jersey

February 2014

EXECUTIVE SUMMARY

To support the decision-making process of education agencies establishing a passing score (cut score) for the Praxis™ Middle School Science (5440) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study.

PARTICIPATING STATES

Panelists from 20 states and Guam were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as either science teachers or college faculty who prepare science teachers and (b) familiarity with the knowledge and skills required of beginning science 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 Middle School Science test, the recommended passing score¹ is 60 out of a possible 100 raw-score points. The scaled score associated with a raw score of 60 is 150 on a 100–200 scale.

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

To support the decision-making process for education agencies establishing a passing score (cut score) for the Praxis™ Middle School Science (5440) test, research staff from ETS designed and conducted a multistate standard-setting study in February 2014 in Princeton, New Jersey. Education agencies² recommended panelists with (a) experience as either science teachers or college faculty who prepare science teachers and (b) familiarity with the knowledge and skills required of beginning science teachers. Twenty states and Guam (Table 1) were represented by 32 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating Jurisdictions and Number of Panelists

Arkansas (2 panelists)	Nevada (1 panelist)
Delaware (1 panelist)	New Hampshire (1 panelist)
Guam (1 panelist)	New Jersey (2 panelists)
Hawaii (1 panelist)	North Carolina (1 panelist)
Idaho (2 panelists)	North Dakota (2 panelists)
Kansas (1 panelist)	Rhode Island (1 panelist)
Kentucky (2 panelists)	South Carolina (2 panelists)
Louisiana (2 panelists)	South Dakota (2 panelists)
Maine (1 panelist)	Virginia (2 panelists)
Maryland (2 panelists)	West Virginia (1 panelist)
Mississippi (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 jurisdiction, the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in accordance with applicable regulations. This study provides a recommended passing score,³ which represents the combined judgments of two panels of experienced educators. Each jurisdiction may want

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

³ In addition to the recommended passing score averaged across the two panels, the recommended passing scores for each panel are presented.

to consider the recommended passing score but also other sources of information when setting the final Praxis Middle School Science passing score (see Geisinger & McCormick, 2010). A jurisdiction 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 jurisdiction'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 Middle School Science test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows a jurisdiction to recognize that any test score on any standardized test—including a Praxis Middle School Science 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 jurisdiction to gauge the likelihood that the recommended passing score from a particular panel would be similar to the passing scores recommended by other panels of experts similar in composition and experience. The smaller the SEJ, the more likely that another panel would recommend a passing score consistent with the recommended passing score. The larger the SEJ, the less likely the recommended passing score would be reproduced by another panel.

In addition to measurement error metrics (e.g., SEM, SEJ), each jurisdiction should consider the likelihood of classification errors. That is, when adjusting a passing score, policymakers should consider whether it is more important to minimize a false-positive decision or to minimize a false-negative decision. A false-positive decision occurs when a candidate's test score suggests that he should receive a license/certificate, but his actual level of knowledge/skills indicates otherwise (i.e., the candidate does not possess the required knowledge/skills). A false-negative decision occurs when a candidate's test score suggests that she should not receive a license/certificate, but she actually does possess the required knowledge/skills. The jurisdiction needs to consider which decision error is more important to minimize.

OVERVIEW OF THE PRAXIS MIDDLE SCHOOL SCIENCE TEST

The Praxis Middle School Science *Test at a Glance* document (ETS, in press) describes the purpose and structure of the test. In brief, the test measures whether entry-level science teachers have the knowledge/skills believed necessary for competent professional practice.

The two and a half-hour test contains 125 selected-response⁴ items covering six content areas: *Scientific Inquiry, Methodology, Techniques, and History* (approximately 15 items), *Basic Principles of Matter and Energy* (approximately 15 items), *Physical Sciences* (approximately 28 items), *Life Sciences* (approximately 30 items), *Earth and Space Sciences* (approximately 22 items), and *Science, Technology, and Society* (approximately 15 items).⁵ The reporting scale for the Praxis Middle School Science ranges from 100 to 200 scaled-score points.

PROCESSES AND METHODS

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

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 reviewed 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.

⁴ Twenty-five of the 125 selected-response items are pretest items and do not contribute to a candidate's score.

⁵ The number of items for each content area may vary slightly from form to form of the test.

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.

DESCRIBING THE JUST QUALIFIED CANDIDATE

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

Both panels worked together to create a description of the just qualified candidate — the knowledge/skills that differentiate a *just* from a *not quite* qualified candidate. To create this description, they first split into smaller groups to consider the just qualified candidate. Then they reconvened and, through whole-group discussion, created the description of the just qualified candidate to use for the remainder of the study. After the description was completed, panelists were split into two, distinct panels that worked separately for the remainder of the study.

The written description of the just qualified candidate summarized the 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 Middle School Science 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 decided if, overall, the item would be

difficult for the just qualified candidate, easy for the just qualified candidate or moderately difficult/easy. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Difficult items for the just qualified candidate are in the 0 to .30 range.
- Moderately difficult/easy items for the just qualified candidate are in the .40 to .60 range.
- Easy items for the just qualified candidate are in the .70 to 1 range.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that an item would be easy for the just qualified candidate, the initial decision located the item in the .70 to 1 range. The second decision for the panelist was to decide if the likelihood of answering it correctly is .70, .80, .90, .95 or 1.

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

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

The panelists discussed their item-level judgments. These discussions helped panelists maintain a shared understanding of the knowledge/skills of the just qualified candidate and helped to clarify aspects of items that might not have been clear to all panelists during the Round 1 judgments. The purpose of the discussion was not to encourage panelists to conform to another's judgment, but to understand the different relevant perspectives among the panelists.

In Round 2, panelists discussed their Round 1 judgments and were encouraged by the facilitator (a) to share the rationales for their judgments and (b) to consider their judgments in light of the rationales provided by the other panelists. Panelists recorded their Round 2 judgments only for items when they wished to change a Round 1 judgment. Panelists' final judgments for the study, therefore, consist of their Round 1 judgments and any adjusted judgments made during Round 2.

Other than the description of the just qualified candidate, results from Panel 1, including the summary of the Round 1 judgments, were not shared with Panel 2. The item-level judgments and resulting discussions for Panel 2 were independent of judgments and discussions that occurred with Panel 1.

RESULTS

EXPERT PANELS

Table 2 presents a summary of the panelists' demographic information. The panel included 32 educators representing 20 states and Guam. (See Appendix A for a listing of panelists.) Seventeen panelists were teachers, eleven were college faculty, one was an administrator or department head, and three held another position. All of the faculty members' job responsibilities included the training of science teachers.

Table D1 (in Appendix D) presents a summary of demographic information by panel.

Table 2

Panel Member Demographics (Across Panels)

	<i>N</i>	<i>%</i>
Current position		
Teacher	17	53%
Administrator/Department head	1	3%
College faculty	11	34%
Other	3	9%
Race		
White	20	63%
Black or African American	5	16%
Hispanic or Latino	2	6%
Asian or Asian American	2	6%
American Indian or Alaskan Native	1	3%
Native Hawaiian or Other Pacific Islander	1	3%
Other	1	3%