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

Student and Teacher Assessments: Validity and Test Use

ETS

Princeton, New Jersey

May 2021

Executive Summary

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

Participating States

Panelists from 15 states, and Washington, D.C., were recommended to serve on one of two panels by their respective education agencies. The education agencies recommended panelists with (a) experience as either chemistry teachers or college faculty who prepare chemistry teachers and (b) familiarity with the knowledge and skills required of beginning chemistry teachers.

Recommended Passing Score

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the *Praxis* Chemistry test, the recommended passing score[1] is 56 out of a possible 100 raw-score points. The scale score associated with a raw score of 56 is 146 on a 100–200 scale.

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

Introduction

To support the decision-making process for education agencies establishing a passing score (cut score) for the *Praxis*® Chemistry (5246) test, research staff from ETS designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012) in May 2021. Education agencies[2] recommended panelists with (a) experience as either chemistry teachers or college faculty who prepare chemistry teachers and (b) familiarity with the knowledge and skills required of beginning chemistry teachers. Fifteen states, and Washington, D.C., (Table 1) were represented by 25 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1
Participating States , and Washington, D.C., and the Number of Panelists

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

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

ETS provides a recommended passing score from the multistate standard-setting study to education agencies. In each state, and D.C., the department of education, the board of education, or a designated educator licensure board is responsible for establishing the operational passing score in accordance with applicable regulations. This study provides a recommended passing score, which represents the combined judgments of a group of experienced educators. Each state, and D.C. may want to consider the recommended passing score but also other sources of information when setting the final

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

Praxis Chemistry passing score (see Geisinger & McCormick, 2010). A state, and D.C. may accept the recommended passing score, adjust the score upward to reflect more stringent expectations, or adjust the score downward to reflect more lenient expectations. There is no *correct* decision; the appropriateness of any adjustment may only be evaluated in terms of its meeting the state, and D.C.'s needs.

Two sources of information to consider when setting the passing score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the *Praxis* Chemistry test score and the latter, the reliability of panelists' passing-score recommendation. The SEM allows states, and D.C. to recognize that any test score on any standardized test—including a *Praxis* Chemistry test score—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the test score to the *true* score? The SEJ allows states, and D.C. to gauge the likelihood that the recommended passing score from the current panel would be similar to the passing scores recommended by other panels of experts similar in composition and experience. The smaller the SEJ, the more likely that another panel would recommend a passing score consistent with the recommended passing score. The larger the SEJ, the less likely the recommended passing score would be reproduced by another panel.

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

Overview of the *Praxis*® Chemistry Test

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

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

Processes and Methods

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

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

Reviewing the Test

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

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

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

Defining the Just-Qualified Candidate

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

Panel 1 created a description of the just-qualified candidate—the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panel first split into smaller groups to consider the just-qualified candidate. Then they reconvened and, through whole-group discussion, determined the description of the just-qualified candidate to use for the remainder of the study.

The written description of the just-qualified candidate summarized the panel discussion in a list format. The description was not intended to describe all the knowledge and skills of the just-qualified candidate but only highlight those that differentiate a *just-qualified candidate* from a *not-quite-qualified* candidate. The written description was distributed to panelists to use during later phases of the study (see Appendix C for the just-qualified candidate description).

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

Panelists' Judgments

The standard-setting process for the *Praxis* Chemistry test was a probability-based Modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006). Using this method, each panelist judged each item on the likelihood (probability or chance) that the just-qualified candidate would answer the item correctly. Panelists made their judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1. The lower the value, the less likely it is that the just-qualified candidate

would answer the item correctly because the item is difficult for the just-qualified candidate. The higher the value, the more likely it is that the just-qualified candidate would answer the item correctly.

Panelists were asked to approach the judgment process in two stages. First, they reviewed both the description of the just-qualified candidate and the item and determined the probability that the just-qualified candidate would answer the question correctly. The facilitator encouraged the panelists to consider the following rules of thumb to guide their decision:

- Items in the 0 to .30 range were those the just-qualified candidate would have a *low chance* of answering correctly.
- Items in the .40 to .60 range were those the just-qualified candidate would have a *moderate chance* of answering correctly.
- Items in the .70 to 1 range were those that the just-qualified candidate would have a *high chance* of answering correctly.

Next, panelists decided how to refine their judgment within the range. For example, if a panelist thought that there was a *high chance* that the just-qualified candidate would answer the question correctly, the initial decision would be in the .70 to 1 range. The second decision for the panelist was to judge if the likelihood of answering it correctly is .70, .80, .90, .95 or 1.

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

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

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

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

Other than the description of the just qualified candidate, results from Panel 1 were not shared with Panel 2. The item-level judgments from Panel 2 were independent of judgments from Panel 1.

Results

Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panel included 25 educators representing 15 states, and D.C. (See Appendix A for a listing of panelists.) Seventeen panelists were teachers, five were college faculty, one was an administrator or department head, one was a director of field experiences, and one was a chemistry teacher/chair of department. Four of the five faculty members' job responsibilities included the training of chemistry teachers. The number of experts by panel and their demographic information are presented in Appendix D (Table D1).

Table 2
Panel Member Demographics (Across Panels)

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

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

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

Standard-Setting Judgments

Table 3 summarizes the mean passing score recommendations after Round 2 from each panel. Table 3 also includes the standard deviation and the standard error of judgment (SEJ) (Brennon, 2002, Tannenbaum & Katz, 2013). The SEJ is one way of estimating the reliability or consistency of a panel's standard-setting judgments. It indicates how likely it would be for several other panels of educators

similar in makeup, experience, and standard-setting training to the current panel to recommend the same passing score on the same form of the test. (For each panel, the panelists' judgments during Rounds 1 and 2, are presented in Appendix D, Tables D2-D4.)

Table 3

Summary of Round 2 Standard-setting Judgments by Panel

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

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

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

In addition to the recommended passing score for each panel, the average passing score across the two panels is provided to help education agencies determine an appropriate passing score. The panels' average passing score recommendation for the *Praxis* Chemistry test is 55.49 (out of a possible 100 raw-score points). The value was rounded to 56 (next highest raw score) to determine the functional recommended passing score. The scale score associated with 56 raw points is 146.

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

Table 4

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

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

Notes. CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 4.99 raw points. The unrounded CSEM value is added to, or subtracted from, the rounded passing-score recommendation. The resulting values are rounded up to the next-highest whole number and then converted to scale scores.

Final Evaluations

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

Panelists were shown the panel's recommended passing score after Round 2 and asked, in the evaluation, (a) how comfortable they are with the recommended passing score and (b) if they think the score was *too high*, *too low*, or *about right*. A summary of the final evaluation results, per panel, is presented in Appendix D (Tables D7 – D14).

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

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

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

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

Summary

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

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine an appropriate operational passing score. For the *Praxis* Chemistry test, the recommended passing score[5] is 56 out of a possible 100 raw-score points. The scale score associated with a raw score of 56 is 146 on a 100–200 scale.

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

References

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

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

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

ETS. (in press). *The Praxis Series®: The Praxis Study Companion: Chemistry (5246)*. Princeton, NJ: Author.

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

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

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

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

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

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

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

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

Appendix A: Panelists' Names & Affiliations

Participating Panelists With Affiliation and State

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

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

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

Appendix B: Agenda

Praxis® Chemistry (5246)

Standard-Setting Study

Day 1 Agenda

Welcome and Introduction

Overview of Standard Setting and the *Praxis* Chemistry Test

Review the *Praxis* Chemistry Test

Discuss the *Praxis* Chemistry Test

Lunch

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

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

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

Collect Materials; End of Day 1

Praxis® Chemistry (5246)

Standard-Setting Study

Day 2 Agenda

Overview of Day 2

Standard Setting Training in the Modified Angoff Method

Round 1 Standard Setting Judgments

Round 1 Feedback and Round 2 Judgments

Lunch

Feedback on Round 2 Recommended Passing Score

Complete Final Evaluation

Collect Materials; End of Study

Appendix C: Just-Qualified Candidate Description

Description of the Just-Qualified Candidate[6]

A just-qualified candidate…

Nature and Impact of Science and Engineering

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

Principles and Models of Matter and Energy

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

Chemical Composition, Bonding, and Structure

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

Chemical Reactions and Periodicity

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

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

Solutions and Acid-Base Chemistry

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

Appendix D: Panel-Specific Results

Table D1

Panel Member Demographics per Panel

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

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

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

Table D2

Panel 1 Passing Score Summary by Round of Judgments

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

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

Table D3

Panel 2 Passing Score Summary by Round of Judgments

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

Table D4 summarizes each panel's judgments, per round. The Round 1 judgments are made without discussion among the panelists. Therefore, the most variability in judgments is typically present in the first round. Round 2 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease—indicating convergence among the panelists' judgments—was observed for both panels.

Table D4

Summary of Standard-setting Judgments by Panel and by Round

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

Tables D5 and D6 show, for Panel 1 and 2, respectively, the raw scores and the scale scores associated with one and two CSEM below and above each panel's recommended passing scores. The unrounded CSEM value is added to, or subtracted from, the rounded passing-score recommendation. The resulting values are rounded up to the next-highest whole number and then converted to scale scores.

Table D5

Scores 1 and 2 CSEM Around the RPS from Panel 1

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

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

Table D6

Scores 1 and 2 CSEM Around the RPS from Panel 2

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

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

Table D7: Panel 1 Final Evaluation: Process Questions

Likert Statement	Strongly agree *N*	Strongly agree %	Agree *N*	Agree %	Disagree *N*	Disagree %	Strongly disagree *N*	Strongly disagree %
I understood the purpose of this study.	13	93	1	7	0	0	0	0
The instructions and explanations provided by the facilitators were clear.	11	79	3	21	0	0	0	0
The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	11	79	3	21	0	0	0	0
The explanation of how the recommended passing score is computed was clear.	11	79	3	21	0	0	0	0
The opportunity for feedback and discussion between rounds was helpful.	10	71	4	29	0	0	0	0
The process of making the standard-setting judgments was easy to follow.	10	71	4	29	0	0	0	0

Table D8: Panel 1 Final Evaluation: Influences in Standard-Setting Judgments

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential *N*	Very influential %	Somewhat influential *N*	Somewhat influential %	Not influential *N*	Not influential %
The description of the just-qualified candidate	13	93	1	7	0	0
The between-round discussions	6	43	8	57	0	0
The knowledge/skills required to answer each test item	13	93	1	7	0	0
The passing scores of other panel members	4	29	10	71	0	0
My own professional experience	12	86	2	14	0	0

Table D9: Panel 1 Final Evaluation: Comfort with the Panel's Recommendation

Question	Very comfort-able *N*	Very comfort-able %	Somewhat comfort-able *N*	Somewhat comfort-able %	Somewhat uncom-fortable *N*	Somewhat uncom-fortable %	Very uncom-fortable *N*	Very uncom-fortable %
Overall, how comfortable are you with the panel's recommended passing score?	10	71	4	29	0	0	0	0

Table D10: Panel 1 Final Evaluation: Opinion of the Final Recommendation

Statement	Too low *N*	Too low %	About right *N*	About right %	Too high *N*	Too high %
Overall, the recommended passing score is:	2	14	12	86	0	0

Table D11: Panel 2 Final Evaluation: Process Questions

Likert Statement	Strongly agree *N*	Strongly agree %	Agree *N*	Agree %	Disagree *N*	Disagree %	Strongly disagree *N*	Strongly disagree %
I understood the purpose of this study.	11	100	0	0	0	0	0	0
The instructions and explanations provided by the facilitators were clear.	11	100	0	0	0	0	0	0
The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	11	100	0	0	0	0	0	0
The explanation of how the recommended passing score is computed was clear.	11	100	0	0	0	0	0	0
The opportunity for feedback and discussion between rounds was helpful.	11	100	0	0	0	0	0	0
The process of making the standard-setting judgments was easy to follow.	11	100	0	0	0	0	0	0

Table D12: Panel 2 Final Evaluation: Influences in Standard-Setting Judgments

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential *N*	Very influential %	Somewhat influential *N*	Somewhat influential %	Not influential *N*	Not influential %
The description of the just-qualified candidate	11	100	0	0	0	0
The between-round discussions	8	73	3	27	0	0
The knowledge/skills required to answer each test item	10	91	1	9	0	0
The passing scores of other panel members	6	55	4	36	1	9
My own professional experience	8	73	3	27	0	0

Table D13: Panel 2 Final Evaluation: Comfort with the Panel's Recommendation

Question	Very comfort-able *N*	Very comfort-able %	Somewhat comfort-able *N*	Somewhat comfort-able %	Somewhat uncom-fortable *N*	Somewhat uncom-fortable %	Very uncom-fortable *N*	Very uncom-fortable %
Overall, how comfortable are you with the panel's recommended passing score?	7	64	4	36	0	0	0	0

Table D14: Panel 2 Final Evaluation: Opinion of the Final Recommendation

Statement	Too low *N*	Too low %	About right *N*	About right %	Too high *N*	Too high %
Overall, the recommended passing score is:	0	0	11	100	0	0

Appendix E: Technical Notes

Standard Error of Judgment (SEJ)

The standard error of judgment (SEJ) is one way of estimating the reliability or consistency of a panel's standard-setting judgments. It indicates how likely it would be for several other panels of educators similar in makeup, experience, and standard-setting training to the current panel to recommend the same threshold score on the same form of the assessment. The SEJ assumes that panelists are randomly selected and that standard-setting judgments are independent. It is seldom the case that panelists are randomly sampled, and only the first round of judgments may be considered independent. The SEJ, therefore, likely underestimates the uncertainty of threshold scores (Tannenbaum & Katz, 2013).

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

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

Outlier Analysis

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

Conditional Standard Error of Measurement (CSEM)

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

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