ets

March 2025

Multistate Standard-Setting Technical Report

ParaPathways (5757):

Reading and Writing (5758) Mathematics (5759)

Table of Contents

Executive Summary

To support the decision-making process of education agencies establishing a passing score (cut score) for the Reading and Writing (5758) and Mathematics (5759) subtests of the *ParaPathways* (5757) test, research staff from Educational Testing Service (ETS) designed and conducted a multistate standard-setting study (Tannenbaum, 2011, 2012).

Participating States

Panelists from five states, were recommended by their respective education agencies. The education agencies recommended panelists with (a) experience as paraeducators, (b) teachers or other staff who train beginning paraeducators and (b) educators who work with paraeducators.

Recommended Passing Score

ETS provides a recommended passing score from the multistate standard-setting study to help education agencies determine appropriate operational passing scores for the two subtests. For the subtests of the *ParaPathways* test, the recommended passing scores are:

- **Reading and Writing (5758):** The recommended passing score is 28 out of a possible 45 raw-score points. The scaled score associated with a raw score of 28 is 332 on a 310–350 scale.
- **Mathematics (5759):** The recommended passing score is 20 out of a possible 30 raw-score points. The scaled score associated with a raw score of 20 is 334 on a 310–350 scale.

Introduction

To support the decision-making process for education agencies establishing passing scores (cut score) for the Reading and Writing (5758) and Mathematics (5759) subtests of the *ParaPathways* (5757) test, research staff from ETS designed and conducted a multistate standard-setting study (Cizek & Bunch, 2007; Tannenbaum, 2011, 2012) in March 2025. Education agencies recommended panelists with (a) experience as paraeducators, (b) teachers or other staff who train beginning paraeducators and (b) educators who work with paraeducators. Five states, (Table 1) were represented by 10 panelists. (See Appendix A for the names and affiliations of the panelists.)

Table 1

Participating States and the Number of Panelists

Arkansas (1 panelist)	Minnesota (1 panelist)
Connecticut (1 panelist)	Maryland (4 panelists)
Delaware (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 recommended passing scores, which represents the combined judgments of a group of experienced educators and paraeducators. Each state may want to consider the recommended passing scores but also other sources of information when setting the Reading and Writing (5758) and Mathematics (5759) subtests of the *ParaPathways* test (see Geisinger & McCormick, 2010; Cizek, 2020). A state may accept one or both recommended passing scores, adjust one or both scores upward to reflect more stringent expectations, or adjust one or both scores downward to reflect more lenient expectations. There are no correct decisions; the appropriateness of any adjustment may only be evaluated in terms of its meeting the state's needs.

Two sources of information to consider when setting the passing scores are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of the subtest scores 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 the subtest scores from the *ParaPathways* test—is not perfectly reliable. A test score only *approximates* what a candidate truly knows or truly can do on the test. The SEM, therefore, addresses the question: How close of an approximation is the

test score to the *true* score? The SEJ allows states to gauge the likelihood that the recommended passing scores 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 scores would be reproduced by another panel.

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

Overview of the Test

The *Study Companion* document (ETS, in press) for the Reading and Writing and Mathematics subtests of the *ParaPathways* test describes the purpose and structure of the subtests. In brief, both subtests measures whether entry-level paraeducators have the knowledge/skills believed necessary for competent professional practice.

The 85 minute Reading and Writing (5758) subtest contains 51 selected-response[1] items covering two content areas: *Reading* (approximately 30 items) and *Writing* (approximately 21 items).[2] The reporting scale ranges from 310–350 scale-score points.

The 60-minute Mathematics (5759) subtest contains 36 selected-response and numeric entry items[3] covering two content areas: *Numbers and Operations* (approximately 18 items) and *Geometry, Measurement, & Data Analysis* (approximately 18 items).[3] The reporting scale ranges from 310–350 scale-score points.

[1] Six of the 51 selected-response items are pretest items and do not contribute to a candidate's score.
[2] The number of items for each content area may vary slightly from form to form of the test.
[3] Six of the 36 selected-response items are pretest items and do not contribute to a candidate's score.

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 each subtest. This review helped familiarize the panelists with the general structure and content of the subtests.

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

Reviewing the Subtests

For each subtest, the standard-setting panelists took each subtest and then discussed the content measured. This discussion helped bring the panelists to a shared understanding of what each of the subtests do and do not cover, which serves to reduce potential judgment errors later in the standard-setting process.

The discussion covered the major content areas being addressed by each subtest. Panelists were asked to remark on any content areas that would be particularly challenging for beginning paraeducators or areas that address content particularly important for beginning paraeducators. The panel began their review with the Reading and Writing subtest, engaged in their discussion, and drafted their descriptions of the just-qualified candidate (described below). Then, this process was repeated for the Mathematics subtest.

Defining the Just-Qualified Candidate

Following the review of each subtest, panelists described the knowledge and skills of the just-qualified candidate, specific to the content that was measured on the subtest. The *just-qualified candidate description* plays a central role in standard setting (Perie, 2008); the goal of the standard-setting process is to identify the test score that aligns with this description.

The panelists created a description of the just-qualified candidate, focusing on the knowledge/skills that differentiate a *just-qualified* from a *not quite-qualified* candidate. To create this description, the panel first split into three breakout groups[4] to consider the just-qualified candidate. in order to create a draft description for each subtest. Then they reconvened and, through whole-group discussion of the three drafts, reached consensus on the final version to use for the remainder of the study.

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

[4] The members in each breakout groups were randomized for their work on each subtest.

Panelists' Judgments

The standard-setting process for the Reading and Writing (5758) and Mathematics (5759) subtests of the *ParaPathways* (5757) test was a probability-based Modified Angoff method (Brandon, 2004; Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006). Using this method, each panelist judged each item on the likelihood (probability or chance) that the just-qualified candidate would answer the item correctly. Panelists made their judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1. The lower the value, the less likely it is that the just-qualified candidate would answer the item correctly because the item is difficult for the just-qualified candidate. The higher the value, the more likely it is that the just-qualified candidate would answer the item correctly.

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

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

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

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

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

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

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

Results

Expert Panels

Table 2 presents a summary of the panelists' demographic information. The panel included 10 educators representing five states. (See Appendix A for a listing of panelists.) Four panelists were paraeducators, three were teachers, one was college faculty, two were educational specialists, and one was an assistant superintendent for pupil services.

Table 2
Panel Member Demographics

Background Survey Question	Number	Percent
What is your current position?	**N**	**%**
Paraeducator	4	40
Teacher	3	30
Educational specialist	2	20
Assistant superintendent for pupil services	1	10
How do you describe yourself (i.e., race/ethnicity)?	**N**	**%**
Black or African American	5	50
Hispanic or Latino	1	10
White	4	40
What is your gender?	**N**	**%**
Female	7	70
Male	3	30
Are you currently working as an educator or paraeducator in your state?	**N**	**%**
Yes	9	90
No	1	10
I am not currently working at the P-12 level	0	0
Are you currently supervising or mentoring paraeducators?	**N**	**%**
Yes	7	70
No	3	30
I am not currently working at the P-12 level	0	0

Background Survey Question	Number	Percent
Including this year, how many years of experience do you have as an educator or paraeducator?	**N**	**%**
3 years or less	1	10
4–7 years	1	10
8–11 years	0	0
12–15 years	1	10
16 years or more	7	70
I am not currently working at the P-12 level	0	0
Which best describes the location of your K–12 school?	**N**	**%**
Urban	4	40
Suburban	2	20
Rural	3	30
Not currently working at the K–12 level	1	10

Standard-Setting Judgments

For each subtest, tables 3 and 4 show the passing score recommendations of each panelist at each round—the number of raw points needed to "pass" the test. The recommendations are the raw score points needed out of a maximum of 45 for the Reading and Writing (5758) subtest and a maximum of 30 for the Mathematics (5759) subtest.

Table 3

Raw Score Recommendation of Each Panelist by Round of Judgments – Reading and Writing (5758)

Panelist	Round 1	Round 2
1	35.30	32.20
2	21.15	23.65
3	21.50	23.00
4	28.80	29.05
5	32.60	31.30
6	21.80	22.10
7	27.65	28.25
8	39.25	28.35
9	33.90	31.70
10	28.20	27.50

Table 4

Raw Score Recommendation of Each Panelist by Round of Judgments – Mathematics (5759)

Panelist	Round 1	Round 2
1	22.40	21.40
2	20.00	21.20
3	19.80	19.90
4	21.50	21.20
5	17.00	17.60
6	19.25	19.15
7	19.10	19.60
8	17.10	18.60
9	19.00	18.70
10	17.95	18.25

Tables 5 and 6 summarize the standard-setting judgments of the panel at each round of judgment for each subtest. The mean represents the panel's passing score recommendation at each round. Tables 5 and 6 also include the standard deviation and the standard error of judgment (SEJ). The SEJ is one way of estimating the reliability or consistency of a panel's standard-setting judgments. It indicates how likely it would be for several other panels of educators similar in makeup, experience, and standard-setting training to the current panel to recommend the same passing score on the same form of the test. (Appendix D provides the technical notes, which further describe the SEJ.)

Table 5

Summary Statistics by Round of Judgments – Reading and Writing (5758)

Statistic	Round 1	Round 2
Mean	29.02	27.71
Minimum	21.15	22.10
Maximum	39.25	32.20
SD	6.27	3.68
SEJ	1.98	1.16

Table 6

Summary Statistics by Round of Judgments – Mathematics (5759)

Statistic	Round 1	Round 2
Mean	19.31	19.56
Minimum	17.00	17.60
Maximum	22.40	21.40
SD	1.74	1.34
SEJ	0.55	0.43

Round 1 judgments are made without discussion among the panelists. Therefore, there is typically more variability in judgments present in the first round. Round 2 judgments, however, are informed by panel discussion; thus, it is common to see a decrease both in the standard deviation and SEJ. This decrease—indicating convergence among the panelists' judgments—was observed (see Tables 5 and 6).

The Round 2 mean score is the panel's final recommended passing score. The panel's passing

score recommendation for the Reading and Writing subtest is 27.71 (out of a possible 45 raw-score points). The value was rounded to the next highest whole number, 28, to determine the functional recommended passing score. The scale score associated with 28 raw points is 332.

The conditional standard error of measurement (CSEM) around the recommended passing score is 3.29 raw points. A standard error represents the uncertainty associated with a test score (See Appendix D for further information about the CSEM.) Table 7 shows the raw scores and the scale scores associated with one and two CSEM below and above the recommended passing score for the Reading and Writing subtest.

Table 7

Scores 1 and 2 CSEM Around the Recommended Passing Score (RPS) – Reading and Writing (5758)

Scores	Raw Score Points out of 45	Scale Score Equivalent
RPS - 2 CSEM	22	324
RPS - 1 CSEM	25	328
RPS	**28**	**332**
RPS +1 CSEM	32	337
RPS +2 CSEM	35	340

Notes. CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 3.29 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.

Table 8 shows that the panel's passing score recommendation (the Round 2 mean score) for the Mathematics subtest is 19 (out of a possible 30 raw-score points). The value was rounded to the next highest whole number, 20, to determine the functional recommended passing score. The scale score associated with 20 raw points is 334.

The conditional standard error of measurement (CSEM) around the recommended passing score is 2.63 raw points. A standard error represents the uncertainty associated with a test score (See Appendix D for further information about the CSEM.) Table 8 shows the raw scores and the scale scores associated with one and two CSEM below and above the recommended passing score for the Mathematics subtest.

Table 8

Scores 1 and 2 CSEM Around the Recommended Passing Score (RPS) – Mathematics (5759)

Scores	Raw Score Points out of 30	Scale Score Equivalent
RPS - 2 CSEM	15	325
RPS - 1 CSEM	18	330
RPS	**20**	**334**
RPS +1 CSEM	23	340
RPS +2 CSEM	26	345

Notes. CSEM = conditional standard error(s) of measurement. The CSEM of the recommended passing score is 2.63 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 scores for both subtests after completing their Mathematics Round 2 judgments and asked, in the evaluation, (a) how comfortable they are with the recommended passing score and (b) if they think the score was *too high*, *too low*, or *about right*. A summary of the final evaluation results is presented in Appendix E.

Nine of the 10 panelists *strongly agreed* or *agreed* that they understood the purpose of the study; one panelist *disagreed*. Nine of the 10 panelists *strongly agreed* or *agreed* that the facilitator's instructions and explanations were clear; one panelist *disagreed*. It is important to note that after the training in the standard-setting methodology and the practice round, which included discussion, all panelists indicated that they understood the purpose of the study and that they were ready to proceed. 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* in guiding their standard-setting judgments. All of the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments. Eight of the 10 panelists indicated that their own professional experience was *very influential* in guiding their judgments.

For the Reading and Writing (5758) subtest, eight of the 10 panelists indicated they were *very comfortable* with the passing score they recommended; two were *somewhat comfortable*. Nine of the 10 panelists indicated the recommended passing score was *about right*; one panelist felt that the recommended passing score was *too high*.

For the Mathematics (5759) subtest, all of the panelists indicated they were *very comfortable* with the passing score they recommended. All of the panelists indicated the recommended passing score was *about right*.

Summary

To support the decision-making process for education agencies establishing passing scores (cut score) for the Reading and Writing (5758) and Mathematics (5759) subtests of the *ParaPathways* 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 subtests of the *ParaPathways* test, the recommended passing scores are:

- **Reading and Writing:** The recommended passing score is 28 out of a possible 45 raw-score points. The scaled score associated with a raw score of 28 is 332 on a 100–200 scale.
- **Mathematics:** The recommended passing score is 20 out of a possible 30 raw-score points. The scaled score associated with a raw score of 20 is 334 on a 100–200 scale.

References

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

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

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

Cizek, G. J. (2020). *Validity: An integrated approach to test score meaning and use*. Routledge.

ETS. (in press). *ParaPathways (5757)*. Princeton, NJ: Author.

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

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

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

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

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

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

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

Appendix A: Panelists' Names & Affiliations

Participating Panelists With Affiliation and State

Panelist Name	Panelists' Affiliation and State Abbreviation
Adebisi Babayemi	Prince George's County Public Schools (MD)
Evanthia Boling	Kennedy Krieger Institute (MD)
Missy Fee	Annandale School District (MN)
Norbert Gouanet	Prince George's County Public Schools (MD)
Cynthia Jones	Love Creek Elementary School (DE)
Theresa May	St Georges Technical High School (DE)
Tilesha McFee	Delaware State Education Association (DE)
Miguel Pabon	Capitol Region Education Council (CT)
Tyshawn Richardson	Kennedy Krieger Institute (MD)
Alisha Ross	Arkansas Department of Education (AR)

Note. All panelists agreed to be listed as participants in the study.

Appendix B: Agenda

***ParaPathways* (5757)**

Standard-Setting Study

Day 1 Agenda	
10:00 AM ET / 9:00 AM CT / 8:00 AM MT / 7:00 AM PT/ 6:00 AM AK / 5:00 AM HI	Welcome, introductions, and goals for the day Standard-setting overview presentation • Q&A about the training • Polling: Checking for understanding Test familiarization for the Reading and Writing (5758) subtest • Overview & instructions • Independent test familiarization • Self-scoring instructions • Independent self-scoring **Break** Discussion of the content measured **Lunch break** Just-Qualified Candidate – Reading and Writing (5758) subtest • Overview • Polling: Evaluation of JQC training • Small group drafts **Break** Test familiarization for the Mathematics (5759) subtest • Independent test familiarization • Self-scoring instructions • Independent self-scoring Discussion of the content measured Just-Qualified Candidate – Mathematics (5759) subtest • Small group drafts
7:00 PM ET / 6:00 PM CT / 5:00 PM MT / 4:00 PM PT / 3:00 PM AK / 2:00 PM HI	End of Day 1

***ParaPathways* (5757)**

Standard-Setting Study

Day 2 Agenda	
10:00 AM ET / 9:00 AM CT / 8:00 AM MT / 7:00 AM PT / 6:00 AM AK / 5:00 AM HI	Overview of the Day Just-Qualified Candidate – Reading and Writing (5758) subtest • Whole group consensus **Break** Just-Qualified Candidate – Mathematics (5759) subtest • Whole group consensus **Lunch break** Standard Setting Training for Selected-Response Items • Instructions and materials • Independent practice round judgments **Break** Practice Round Data Discussion • Instructions, materials, and screen setup • Discussion of the practice round data • Polling: Evaluation of the judgment training Round 1 Standard Setting Judgments – Reading and Writing (5758) subtest
7:00 PM ET/ 6:00 PM CT / 5:00 PM MT / 4:00 PM PT / 3:00 PM AK / 2:00 PM HI	End of Day 2

***ParaPathways* (5757)**

Standard-Setting Study

Day 3 Agenda	
10:00 AM ET / 9:00 AM CT / 8:00 AM MT / 7:00 AM PT/ 6:00 AM AK / 5:00 AM HI	Overview of the Day Honoraria Payment Process Presentation Round 1 feedback: Summary data – Reading and Writing (5758) subtest • Polling: Evaluation of data presentation & Round 2 training Round 1 feedback: Item-level data and Round 2 judgments (**break** as needed) – Reading and Writing (5758) subtest **Lunch break** Round 1 Standard Setting Judgments – Mathematics (5759) subtest **Break** Round 1 feedback: Summary data – Mathematics (5759) subtest Round 1 feedback: Item-level data and Round 2 judgments (break as needed) – Mathematics (5759) subtest **Break** Round 2 feedback: Recommended Passing Scores for both subtests Complete final evaluation Wrap Up/ Final Steps • Review Nondisclosure of Secure Materials o Destruction of materials o What you can/cannot discuss • Final Questions and Thank You
7:00 PM ET/ 6:00 PM CT / 5:00 PM MT / 4:00 PM PT / 3:00 PM AK / 2:00 PM HI	End of Study

Appendix C: Just-Qualified Candidate Descriptions

Description of the Just-Qualified Candidate[5]

Reading & Writing (5758)

A just-qualified candidate…

1. Can infer information that is directly stated in informational text
2. Can identify main ideas, themes, and supporting details of text with minimum complexity
3. Can identify author's purpose and point of view of a text with minimum complexity.
4. Can use context clues to understand unfamiliar vocabulary when reading simple text.
5. Can identify evidence to support an argument within a simple text or simple visual representation
6. Can clearly organize ideas and make connections with supporting details.
7. Can revise simple text for clarity, grammar, conventions, and usage for comprehension
8. Identifies relevant information about a topic, recognizes credible sources, and attempts to give credit to the sources
9. Can Identify and apply basic principles of phonics
10. Can communicate with students in different ways to support them with their writing (e.g., revising & revisiting) and reading comprehension (e.g., through questioning)

[5] 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[6]

Mathematics (5759)

A just-qualified candidate…

1. Uses the four operations to solve problems involving integers.
2. Solves one and two-step problems using decimals, percents, and fractions
3. Can compare two quantities
4. Develops equations to solve basic word problems.
5. Solves simple Geometrical problems involving area, perimeter, volume
6. Solves one-step ratios and proportions problems using pie-charts, tables, or other visuals.
7. Solves single variable linear equations.
8. Uses properties of operations to simplify algebraic expressions.
9. Understands the defining attributes of two-dimensional shapes and can use them to solve geometric problems.
10. Has the ability to convert within the same measurement system, given the rule
11. Understands of order of operations.

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

Appendix D: Technical Notes

Standard Error of Judgment (SEJ)

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

The SEJ (SEJ_{Mn}) 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).

$$SEJ_{\mathrm{Mn}} = \mathrm{SD}/\sqrt{n}$$

Conditional Standard Error of Measurement (CSEM)

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

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

Appendix E: Final Evaluation Results

Table E1

Final Evaluation Process Questions

Likert Statement	Strongly agree *N*	Strongly agree %	Agree *N*	Agree %	Disagree *N*	Disagree %	Strongly disagree *N*	Strongly disagree %
I understood the purpose of this study.	7	70	2	20	1	10	0	0
The instructions and explanations provided by the facilitator were clear.	6	60	3	30	1	10	0	0
The training in the standard-setting method was adequate to give me the information I needed to complete my assignment.	7	70	3	30	0	0	0	0
The explanation of how the recommended passing score is computed was clear.	6	40	4	40	0	0	0	0
The opportunity for feedback and discussion for round 2 judgments was helpful.	9	90	1	10	0	0	0	0
The process of making the standard-setting judgments was easy to follow.	4	40	6	60	0	0	0	0

Table E2

Final Evaluation: Standard-Setting Process

	Too much time *N*	Too much time %	About the right amount of time *N*	About the right amount of time %	Too little time *N*	Too little time %
Small group JQC drafts	0	0	9	90	1	10
Whole group JQC consensus	1	10	9	90	0	0
Training and practice for making standard-setting judgments	1	10	9	90	0	0
Round 1 judgments (independent)	2	20	8	80	0	0
Round 2 judgments (with discussion)	1	10	9	90	0	0

Table E3

Final Evaluation: Influences in Standard-Setting Judgments

How influential was each of the following factors in guiding your standard-setting judgments?	Very influential *N*	Very influential %	Somewhat influential *N*	Somewhat influential %	Not influential *N*	Not influential %
The description of the just-qualified candidate	10	100	0	0	0	0
The between-round discussions	7	70	3	30	0	0
The knowledge/skills required to answer each test item	8	80	2	20	0	0
The passing scores of other panel members	4	40	6	60	0	0
My own professional experience	8	80	2	20	0	0

Table E4

Final Evaluation: Comfort with the Panel's Recommendation

Question	Very comfort-able *N*	Very comfort-able %	Somewhat comfort-able *N*	Somewhat comfort-able %	Somewhat uncom-fortable *N*	Somewhat uncom-fortable %	Very uncom-fortable *N*	Very uncom-fortable %
Overall, how comfortable are you with the panel's recommended passing score?								
Reading and Writing	8	80	2	20	0	0	0	0
Mathematics	10	100	0	0	0	0	0	0

Table E5

Final Evaluation: Opinion of the Final Recommendation

Statement	Too low *N*	Too low %	About right *N*	About right %	Too high *N*	Too high %
Overall, the recommended passing score is:						
Reading and Writing	0	0	9	90	1	10
Mathematics	0	0	10	100	0	0