

Board of Education Agenda Item

Item: _____ N. _____

Date: April 28, 2011

Topic: First Review of a Recommendation of the Advisory Board on Teacher Education and Licensure to Approve a Cut Score for the Reading for Virginia Educators: Reading Specialist Assessment

Presenter: Mrs. Patty S. Pitts, Assistant Superintendent for Teacher Education and Licensure

Telephone Number: 804-371-2522

E-Mail Address: Patty.Pitts@doe.virginia.gov

Origin:

Topic presented for information only (no board action required)

Board review required by
 State or federal law or regulation
 Board of Education regulation
 Other: _____

Action requested at this meeting Action requested at future meeting: May 19, 2011 (date)

Previous Review/Action:

No previous board review/action

Previous review/action
date _____
action _____

Background Information:

In response to House Joint Resolution Number 794 (HJR 794) of the 2001 session of the Virginia General Assembly, the Advisory Board on Teacher Education and Licensure (ABTEL), in cooperation with the State Council of Higher Education for Virginia, conducted a series of initiatives to determine the proficiency of Virginia teachers in teaching systematic explicit phonics. A resolution to enhance reading instruction was adopted on March 17, 2003, by ABTEL. The resolution was presented to the Board of Education for first review on March 26, 2003, and approved by the Board on April 29, 2003. This resolution called for the following:

1. the development of a statewide reading assessment aligned with the Virginia Standards of Learning and the National Reading Panel's five key components of effective reading instruction: phonics, phonemic awareness, vocabulary, comprehension, and fluency; and

2. the requirement of a reading instructional assessment for teachers of special education (Emotional Disturbances, Learning Disabilities, Mental Retardation, Hearing Impairments, and Visual Impairments), elementary prek-3, and elementary prek-6 no later than July 1, 2004. In addition, individuals seeking a reading specialist endorsement would be required to complete a reading instructional assessment no later than July 1, 2004.

In response to this resolution, the Virginia Department of Education contracted with National Evaluation Systems to develop the Virginia Reading Assessment (VRA) and Virginia Reading Assessment for Reading Specialists (VRA for Reading Specialists). Between July 1, 2004, and June 30, 2006, the VRA was required of all candidates applying for an initial license with endorsements in Early/Primary PreK-3, Elementary Education PreK-6, Special Education (Emotional Disturbances, Learning Disabilities, Mental Retardation, Hearing Impairments, and Visual Impairments) and individuals seeking an endorsement as a Reading Specialist. Also, as a result of the Board's action on July 27, 2005, institutions of higher education with preparation programs in teaching endorsement areas requiring the VRA were given another year to continue aligning their programs with required reading competencies.

At the July 27, 2005, meeting, the Board of Education approved cut scores for the Virginia Reading Assessments (VRA) for elementary and special education teachers (Emotional Disturbances, Learning Disabilities, Mental Retardation, Hearing Impairments, and Visual Impairments) and reading specialists. The Board approved a score of 235 for elementary and special education teachers and a score of 245 for reading specialists, effective July 1, 2006.

Based on Virginia's procurement regulations, from time to time contracts for certain tests must be opened for competitive solicitation and new contracts awarded. As a result of the solicitation, the Virginia Department of Education contracted with the Educational Testing Service (ETS) on July 20, 2010, to develop the following two new reading assessments that will become effective July 1, 2011.

Reading for Virginia Educators (RVE): Elementary and Special Education Teachers

This assessment will be required for Virginia teachers seeking an initial license with an endorsement in Elementary Education PK-3, Elementary Education PK-6, Special Education-General Curriculum, Special Education-Hearing Impairments, and Special Education-Visual Impairments and will replace the Virginia Reading Assessment (VRA) for Elementary and Special Education Teachers.

Reading for Virginia Educators: Reading Specialist (RVE: Reading Specialist)

This assessment will be required for individuals seeking the reading specialist endorsement and will replace the Virginia Reading Assessment (VRA) for Reading Specialists.

The Educational Testing Service worked with the Virginia Department of Education to assemble test development committees composed of Virginia teachers and higher education faculty involved in the preparation of reading teachers. These committees met in September 2010 to review the proposed test specifications and approve specific test items for the new assessments. ETS also conducted field tests of the two new assessments across Virginia in January and February 2011.

Summary of Major Elements

To support the decision-making process for the Virginia Department of Education with regards to establishing a passing score, or cut score, for the RVE: Reading Specialist (0304) assessment, research staff from Educational Testing Service (ETS) designed and conducted a standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level reading specialists.

The study involved an expert panel comprised of teachers, administrators and college faculty. The Department of Education recommended panelists with (a) reading specialist experience, either as reading specialists or college faculty who prepare reading specialists and (b) familiarity with the knowledge and skills required of beginning reading specialists. A roster of participants is included in the Appendix of the attached report. The panel was convened on February 28 and March 1, 2011, in Richmond, Virginia.

The *RVE: Reading Specialist Test at a Glance* document (ETS, in press) describes the purpose and structure of the assessment. In brief, the assessment measures whether entry-level reading specialists have the content knowledge and skills believed necessary for competent professional practice. The specifications for the assessment were provided by the Virginia Department of Education and consistent with the current knowledge and skill content specified for licensure.

The three and one-half hour assessment is divided into two parts. Part A contains 100 multiple-choice questions covering *Assessment and Diagnostic Teaching* (approximately 18 questions), *Oral Language and Oral Communication* (approximately 12 questions), *Reading Development* (approximately 40 questions), *Writing and Research* (approximately 12 questions) and *Specialized Knowledge and Leadership Skills* (approximately 18 questions)¹. Part B contains a constructed-response question and a case study covering the same five content areas as Part A. While the sections are not separately timed, suggested time limits of 120 minutes for Part A, 30 minutes for the constructed-response question, and 60 minutes for the case study are provided.

Candidate scores on the two parts are combined and reported as an overall score; six category scores – one for each content area covered in Part A and one for the combined constructed-response question and case study in Part B – also are reported. The constructed-response question and case study in Part B are weighted to contribute 25% of the total raw-score points². The maximum total number of raw points that may be earned on the assessment is 107, 80 points from Part A and 27 points from Part B. The reporting scales for the RVE: Reading Specialist (0304) assessment ranges from 100 to 200 scaled-score points.

The panel's cut score recommendation for the RVE: Reading Specialist (0304) assessment is 70.13. The value was rounded to 71, the next highest whole number, to determine the functional recommended cut. The value of 71 represents approximately 66% of the total available 107 raw-score points that could be earned on the assessment. The scaled score associated with 71 raw points is 162 (on a 100 to 200 scale).

When reviewing the Standard Error of Measurement (SEM) for the cut scores recommended by the Virginia Standard Setting Study, there is an overlap in the scaled scores. The SEM is a statistical phenomenon and is unrelated to the accuracy of scoring. All test results are subject to the standard error of measurement. If a test-taker were to take the same test repeatedly, with no change in his level of knowledge and preparation, it is possible that some of the resulting scores would be slightly higher or slightly lower than the score that precisely reflects the test taker's actual level of knowledge and ability. The difference between a test-taker's actual score and his highest or lowest hypothetical score is known

as the standard error of measurement. The Standard Error of Measurement for the recommended cut scores for the Virginia Standard Setting Study is shown below. Note that consistent with the recommended cut score, the cut scores at the different SEMs have been rounded to the next highest whole number.

Standard Error of Measure Summary – Reading for Virginia Educators: Reading Specialist

Cut Scores Within 1 and 2 SEMs of the Recommended Cut Score

<u>Recommended Cut Score (SEM)</u>	<u>Scale Score Equivalent</u>	<u>Field Test Pass Rates</u>	
	71 (4.69)	162 (Panel Recommendation) (ABTEL Recommendation)	70%
-2 SEMs	62	151	79%
-1 SEM	67	158	75%
+1 SEM	76	169	55%
+2 SEMs	81	175	38%

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

In addition to the results of the Standard Setting Study, The Advisory Board on Teacher Education and Licensure (ABTEL) also reviewed the results from the field test conducted by ETS. A total of 164 candidates participated in the field test for the RVE: Reading Specialist assessment conducted in January-February 2011. The percentage of field test candidates passing at the scale score equivalent is also shown above.

On March 21, 2011, ABTEL recommended that the Board of Education set a cut score of 162 for the Reading for Virginia Educators: Reading Specialist assessment.

Superintendent's Recommendation:

The Superintendent of Public Instruction recommends that the Board of Education receive for first review the Advisory Board on Teacher Education and Licensure's recommendation to approve the cut score of 162 for the Reading for Virginia Educators: Reading Specialist assessment.

Impact on Resources:

Costs associated with the administration of the Reading for Virginia Educators: Reading Specialist assessment will be incurred by the Educational Testing Service. Prospective elementary and special education teachers will be required to pay a fee for test administration and reporting results to the Virginia Department of Education.

Timetable for Further Review/Action:

This agenda item will be presented to the Board of Education for final approval at the May 19, 2011, meeting.



Listening. Learning. Leading.

Standard Setting Technical Report

READING FOR VIRGINIA EDUCATORS: READING SPECIALIST (0304)

Prepared for the Virginia Department of Education

Educational and Credentialing Research

Educational Testing Service

Princeton, New Jersey

March 2011

Executive Summary

To support the decision-making process for the Virginia Department of Education (VDOE) with regards to establishing a passing score, or cut score, for the Reading for Virginia Educators (RVE): Reading Specialist (0304) assessment, research staff from Educational Testing Service (ETS) designed and conducted a standard-setting study on February 28 and March 1, 2011. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level reading specialists.

Recommended Cut Score

The recommended cut score is provided to help the VDOE determine an appropriate cut (or passing) score. For the RVE: Reading Specialist (0304) assessment, the average recommended cut score is 71 (on the raw score metric), which represents 66% of total available 107 raw score points. The scaled score associated with a raw score of 71 is 162 (on a 100 to 200 scale).

Summary of Content Specification Judgments

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

To support the decision-making process for the Virginia Department of Education (VDOE) with regards to establishing a passing score, or cut score, for the RVE: Reading Specialist (0304) assessment, research staff from Educational Testing Service (ETS) designed and conducted a standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level reading specialists.

The study involved an expert panel, comprised of teachers, administrators and college faculty. The VDOE recommended panelists with (a) reading specialist experience, either as reading specialists or college faculty who prepare reading specialists and (b) familiarity with the knowledge and skills required of beginning reading specialists.

The panel was convened on February 28 and March 1, 2011, in Richmond, Virginia. The following technical report is divided into three sections. The first section describes the content and format of the assessment. The second section describes the standard-setting processes and methods used. The third section presents the results of the standard-setting study.

The passing score recommendation for the RVE: Reading Specialist (0304) assessment is provided to the VDOE. The VDOE is responsible for establishing the final passing score in accordance with applicable state regulations. The study provides a recommended passing score, which represents the combined judgments of one group of experienced educators. The full range of the VDOE's needs and expectations could not be represented during the standard-setting study. The VDOE, therefore, may want to consider both the panel's recommended cut score and other sources of information when setting the final RVE: Reading Specialist (0304) cut score (see Geisinger & McCormick, 2010). The VDOE may accept the recommended cut score, adjust it upward to reflect more stringent expectations, or adjust it 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 VDOE's needs.

Two sources of information to consider when setting the cut score are the standard error of measurement (SEM) and the standard error of judgment (SEJ). The former addresses the reliability of RVE: Reading Specialist (0304) scores and the latter the reliability of panelists' cut score recommendations. The SEM allows the VDOE to recognize that a RVE: Reading Specialist (0304) score—any test score on any test—is less than perfectly reliable. A test score only approximates what a candidate *truly* knows or *truly* can do on the test. The SEM, therefore, addresses the question: How close

of an approximation is the test score to the *true* score? The SEJ allows the VDOE to consider the likelihood that the recommended cut score from the current panel would be similar to cut 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 cut score consistent with the recommended cut score. The larger the SEJ, the less likely the recommended cut score would be reproduced by another panel.

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

Overview of the RVE: Reading Specialist Assessment

The *RVE: Reading Specialist Test at a Glance* document (ETS, in press) describes the purpose and structure of the assessment. In brief, the assessment measures whether entry-level reading specialists have the content knowledge and skills believed necessary for competent professional practice. The specifications for the assessment were provided by the Virginia Department of Education and consistent with the current knowledge and skill content specified for licensure.

The three and one-half hour assessment is divided into two parts. Part A contains 100 multiple-choice questions covering *Assessment and Diagnostic Teaching* (approximately 18 questions), *Oral Language and Oral Communication* (approximately 12 questions), *Reading Development* (approximately 40 questions), *Writing and Research* (approximately 12 questions) and *Specialized Knowledge and Leadership Skills* (approximately 18 questions)¹. Part B contains a constructed-response question and a case study covering the same five content areas as Part A. While the sections are not

¹ The number of questions for each content area may vary slightly from form to form of the assessment.

separately timed, suggested time limits of 120 minutes for Part A, 30 minutes for the constructed-response question, and 60 minutes for the case study are provided.

Candidate scores on the two parts are combined and reported as an overall score; six category scores – one for each content area covered in Part A and one for the combined constructed-response question and case study in Part B – also are reported. The constructed-response question and case study in Part B are weighted to contribute 25% of the total raw-score points². The maximum total number of raw points that may be earned on the assessment is 107, 80 points from Part A and 27 points from Part B. The reporting scales for the RVE: Reading Specialist (0304) assessment ranges from 100 to 200 scaled-score points.

Processes and Methods

The following section describes the processes and methods used to train panelists, gather panelists' judgments and to calculate the recommended passing score, or cut score. (The agenda for the panel meeting is presented in the Appendix.)

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

The standard-setting study began with a welcome and introduction by James Lanham, from the VDOE. The ETS facilitator, Clyde Reese, then explained how the assessment was developed, provided an overview of standard setting, and presented the agenda for the study.

Reviewing the Assessment

The first activity was for the panelists to “take the test.” (Each panelist had signed a nondisclosure form.) The panelists were given approximately two hours to respond to the multiple-choice questions and to sketch responses to the constructed-response question and case study. (Panelists were instructed not to refer to the answer key for the multiple-choice questions while taking the test.) The purpose of “taking the test” was for the panelists to become familiar with the test format, content,

² The constructed-response question is weighted by a factor of 1.8 (maximum score of 10.8) and the case study is weighted by a factor 2.7 (maximum score of 16.2).

and difficulty. After “taking the test,” the panelists checked their responses against the answer key for the multiple-choice questions and the scoring rubric for the constructed-response question and case study.

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

Defining the Just Qualified Candidate

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

The panelists were split into smaller groups, and each group was asked to write down their definition of a JQC. Each group referred to *RVE: Reading Specialist Test at a Glance* to guide their definition. Each group posted its definition on chart paper, and a full-panel discussion occurred to reach consensus on a final definition (see the consensus JQC definition in the Appendix).

Panelists’ Judgments

The standard-setting process for the RVE: Reading Specialist (0304) assessment was conducted for the overall test, though one standard-setting approach was implemented for Part A (multiple-choice questions) and another approach was implemented for Part B (constructed-response question and case study). The panel’s passing score for the assessment is the sum of the interim cut scores recommended by the panelists for each section. As with scoring and reporting, the panelists’ judgments for Part B, the constructed-response question and case study, were weighted such that Part B contributed 25% of the overall score.

Standard Setting for Part A (multiple-choice questions). A probability-based Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) was used for Part A (multiple-choice questions). In this approach, for each question, a panelist decides on the likelihood (probability or chance) that a JQC

would answer it correctly. Panelists made their judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1. The lower the value, the less likely it is that a JQC would answer the question correctly, because the question is difficult for the JQC. The higher the value, the more likely it is that a JQC would answer the question correctly.

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

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

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

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

Standard Setting for Part B (constructed-response question and case study). An Extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995) was used for Part B (constructed-response question and case study). In this approach, for each question, a panelist decides on the assigned score value that would most likely be earned by a JQC. The basic process that each panelist followed

was first to review the definition of the JQC and then to review the question and the rubric for that question. The rubric for a question defines holistically the quality of the evidence that would merit a response earning a score of 3, 2, 1, or 0. During this review, each panelist independently considered the level of knowledge and/or skill required to respond to the question and the features of a response that would earn 3, 2, 1, or 0 points, as defined by the rubric.

A test taker's response to a constructed-response question is independently scored by two raters, and the sum of the raters' scores is the assigned score³; possible scores, therefore, range from zero (both raters assigned a score of zero) to six (both raters assigned a score of three). Each panelist decided on the score most likely to be earned by a JQC from the following possible values: 0, 1, 2, 3, 4, 5, 6. For the constructed-response question and case study, panelists recorded the score (0 through 6) that a JQC would most likely earn. The panelists practiced making their standard-setting judgments on the constructed-response question in Part B.

Consistent with the standard-setting process used for Part A, the panelists engaged in two rounds of judgments for Part B. Following Round 1, question-level feedback was provided to the panel. The panelists' judgments were displayed for each question. The panelists participated in a general discussion of the results. Panelists were asked to share their rationales for the judgments they made. Following this discussion, panelists were provided an opportunity to change their question-level standard-setting judgments (Round 2).

Judgment of Content Specifications

In addition to the two-round standard-setting process, the panel judged the importance of the knowledge and/or skills stated or implied in the assessment content specifications for the job of an entry-level reading specialist. These judgments addressed the perceived content-based validity of the assessment. Judgments were made using a four-point Likert scale — *Very Important*, *Important*, *Slightly Important*, and *Not Important*. Each panelist independently judged the knowledge/skills statements.

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

Results

Expert Panel

The panel included 16 educators. In brief, 11 panelists were reading specialists⁴, one was an administrator, three were college faculty, and one was a instructional coach. All three of the panelists who were college faculty were currently involved in the training or preparation of reading specialists. Twelve panelists were White, three were African American, and one indicated “other.” Fourteen panelists were female. Fourteen panelists reported being certified reading specialists in Virginia. The majority of panelists (11 of the 16 panelists or 69%) had 11 or fewer years of experience as a reading specialist, and the remainin panelsists (5 of the 16 panelists or 31%) had 16 or more years of experience.

A fuller demographic description for the members of the panel is presented in Table 1. (See Figure 1 in the Appendix for a listing of panelists.)

Table 1

Panel Member Demographics

	N	Percent
Current Position		
Reading Specialist	11	69%
Administrator/Department Head	1	6%
College Faculty	3	19%
Instructional Coach	1	6%
Race		
White	12	75%
Black or African American	3	19%
Other	1	6%
Gender		
Female	14	88%
Male	2	12%
Are you currently certified as a reading specialist in Virginia?		
Yes	14	88%
No	2	12%

⁴ All but one of the 11 panelists who reported their current position as “reading specialist” also reported currently working at the elementary school level. While the reading specialist endorsement in Virginia is K-12, the VDOE indicated that the overwhelming majority of reading specialists work in elementary school settings; therefore, the composition of the panel is representative.

Table 1 (continued)**Panel Member Demographics**

	N	Percent
Are you currently a reading specialist in Virginia?		
Yes	13	81%
No	3	19%
Are you currently supervising or mentoring other reading specialists?		
Yes	9	56%
No	7	44%
How many years of experience do you have as a reading specialist?		
3 years or less	1	6%
4 - 7 years	6	38%
8 - 11 years	4	25%
12 - 15 years	0	0%
16 years or more	5	31%
At what K-12 grade level are you currently working as a reading specialist?		
Elementary (K-5 or K-6)	11	69%
Middle School (6-8 or 7-9)	0	0%
Elementary and Middle School (9 - 12 or 10 - 12)	1	6%
Not currently teaching at the K-12 level	4	25%
Which best describes the location of your K-12 school?		
Urban	4	25%
Suburban	5	31%
Rural	4	25%
Not currently teaching at the K-12 level	3	19%
If you are college faculty, are you currently involved in the training/preparation of reading specialists?		
Yes	3	19%
No	0	0%
Not college faculty	13	81%

Initial Evaluation Forms

The panelists completed an initial evaluation after receiving training on how to make question-level judgments. The primary information collected from this form was the panelists indicating if they

had received adequate training to make their standard-setting judgments and were ready to proceed. All panelists indicated that they were prepared to make their judgments.

Summary of Standard Setting Judgments

A summary of each round of standard-setting judgments for Part A (multiple-choice questions), Part B (constructed-response questions), and the overall assessment is presented in Table 2. The numbers in the table reflect the recommended cut scores — the number of raw points needed to “pass” the part or assessment — of each panelist for the two rounds. For Part B, weighted cut scores are presented; for the overall assessment, the weighted cut scores (i.e., sum of Part A and the weighted Part B cut scores) are presented. Note that the RVE: Reading Specialist (0304) assessment reports a single, overall score and that the panel is recommending a single cut score for the combination of Parts A and B. The separate “cut scores” for the two parts are intermediate steps in calculating the overall cut score. The panel’s average recommended cut score and highest and lowest cut scores are reported, as are the standard deviations (SD) of panelists’ cut scores and the standard errors of judgment (SEJ). The SEJ is one way of estimating the reliability of the judgments⁵. It indicates how likely it would be for other panels of educators similar in make-up, experience, and standard-setting training to the current panels to recommend the same cut score on the same form of the assessment. A comparable panel’s cut score would be within 1 SEJ of the current average cut score 68 percent of the time.

The panel’s cut score recommendation for the RVE: Reading Specialist (0304) assessment is 70.13 (see Table 2). The value was rounded to 71, the next highest whole number, to determine the functional recommended cut. The value of 71 represents approximately 66% of the total available 107 raw-score points that could be earned on the assessment. The scaled score associated with 71 raw points is 162 (on a 100 to 200 scale).

⁵ An SEJ assumes that panel members are randomly selected and that standard-setting judgments are independent. It is seldom the case that panel members are randomly sampled, and only the first round of judgments may be considered independent. The SEJ, therefore, likely underestimates the uncertainty of cut scores (Tannenbaum & Katz, forthcoming).

Table 2**Cut Score Summary by Round of Judgments**

Panelist	Round 1			Round 2		
	Part A	Part B (weighted)	Total	Part A	Part B (weighted)	Total
1	43.50	13.50	57.00	45.35	16.20	61.55
2	55.25	18.00	73.25	55.15	18.00	73.15
3	29.50	13.50	43.00	32.75	16.20	48.95
4	54.00	18.90	72.90	55.20	16.20	71.40
5	62.35	19.80	82.15	61.10	19.80	80.90
6	48.65	15.30	63.95	49.65	15.30	64.95
7	56.65	13.50	70.15	57.15	13.50	70.65
8	50.75	18.00	68.75	50.55	18.00	68.55
9	53.30	18.00	71.30	54.00	18.00	72.00
10	57.75	15.30	73.05	57.85	15.30	73.15
11	54.95	15.30	70.25	55.50	18.00	73.50
12	45.10	15.30	60.40	46.65	15.30	61.95
13	58.70	18.00	76.70	59.05	18.00	77.05
14	55.35	17.10	72.45	54.95	17.10	72.05
15	61.65	15.30	76.95	60.95	18.00	78.95
16	54.20	15.30	69.50	55.40	18.00	73.40
Average	52.60	16.26	68.86	53.20	16.93	70.13
SD	8.06	2.01	9.25	7.11	1.58	7.79
SEJ	2.01	0.50	2.31	1.78	0.40	1.95
Highest	62.35	19.80	82.15	61.10	19.80	80.90
Lowest	29.50	13.50	43.00	32.75	13.50	48.95

Table 3 presents the estimated standard errors of measurement (SEM) around the recommended cut score. A standard error represents the uncertainty associated with a test score. The scaled scores associated with 1 and 2 SEMs above and below the recommended cut scores are provided. The standard errors provided are an estimate, given that the RVE: Reading Specialist (0304) assessment has not yet been administered operationally.

Table 3

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

Recommended Cut Score (SEM)	Scale Score Equivalent
71 (4.69)	162
- 2 SEMs	151
-1 SEM	158
+1 SEM	169
+ 2 SEMs	175

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

Summary of Content Specification Judgments.

Panelists judged the extent to which the knowledge and/or skills reflected by the RVE: Reading Specialist (0304) assessment content specifications were important for entry-level reading specialists. Panelists rated the 18 knowledge/skills statements on a four-point scale ranging from *Very Important* to *Not Important*. All of the knowledge statements were judged to be *Very Important* or *Important* by at least 80% of the panelists. The panelists' ratings are summarized in Table 4 (in Appendix).

Summary of Final Evaluations.

The panelists completed an evaluation form at the conclusion of their standard-setting study. The evaluation form asked the panelists to provide feedback about the quality of the standard-setting implementation. Table 5 (in Appendix) present the results of the final evaluations.

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

The majority of panelists (14 of 16 or 88%) reported that the definition of the JQC was *very influential* in guiding their standard-setting judgments; all reported that it was at least *somewhat influential*. All the panelists reported that between-round discussions were at least *somewhat influential* in guiding their judgments.

All of the panelists indicated they were *very* or *somewhat comfortable* with their recommendation. Approximately 80% of the panelists indicated that the recommend cut score was *about right* (13 of the 16 panelists). The remaining panelists indicated the cut score was *too low*.

Summary

To support the decision-making process for the Virginia Department of Education (VDOE) with regards to establishing passing score, or cut score, for RVE: Reading Specialist (0304) assessment, research staff from Educational Testing Service (ETS) designed and conducted a standard-setting study. The study also collected content-related validity evidence to confirm the importance of the content specifications for entry-level reading specialists.

The recommended cut score is provided to help the VDOE determine an appropriate cut (or passing) score. For RVE: Reading Specialist (0304), the average recommended cut score is 71 (on the raw score metric), which represents 66% of total available 107 raw score points. The scaled score associated with a raw score of 71 is 162.

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

References

- Brandon, P.R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education, 17*, 59-88.
- Cizek, G. J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- ETS. (In press). *Reading for Virginia Educators: Reading Specialist Test At A Glance*. Princeton, NJ.
- Hambleton, R. K., & Pitoniak, M.J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433-470). Westport, CT: American Council on Education/Praeger.
- Hambleton, R. K., & Plake, B.S. (1995). Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41-55.
- Tannenbaum, R.J., & Katz, I.R. (forthcoming). Standard setting. In K.F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology*. Washington, DC: American Psychological Association.

Appendix

AGENDA

Reading for Virginia Educators: Reading Specialist (0304) Standard-setting study

February 28, 2011

- | | |
|---------------|--|
| 8:00 – 8:30 | Welcome and Introduction <ul style="list-style-type: none">• Welcome by the Virginia Department of Education• Overview of Workshop Events |
| 8:30 – 8:45 | Overview of Standard Setting & Workshop Events |
| 8:45 – 9:10 | Overview of the RVE: Reading Specialist Assessment |
| 9:10 – 9:15 | Break |
| 9:15 – 11:30 | “Take” the RVE Reading Specialist Assessment |
| 11:30 – 12:00 | Discuss the RVE Reading Specialist Assessment |
| 12:00 – 12:45 | Lunch |
| 12:45 – 3:00 | Define the Knowledge/Skills of a JQC |
| 3:00 – 3:05 | Break |
| 3:05 – 3:30 | Standard Setting Training for M-C Items |
| 3:30 – 5:15 | Round 1 Standard Setting Judgments for Multiple-Choice <ul style="list-style-type: none">• Questions 1-60 |
| 5:15 – 5:30 | Collect Materials; End of Day 1 |

AGENDA

Reading for Virginia Educators: Reading Specialist (0304) Standard-setting study

March 1, 2011

9:00 – 9:15	Overview of Day 2
9:15 – 10:00	Round 1 Standard Setting Judgments for Multiple-Choice <ul style="list-style-type: none">• Questions 61-100
10:00 – 10:30	Standard Setting Training for CR Items
10:30 – 11:00	Round 1 Standard Setting Judgments for Constructed-Response <ul style="list-style-type: none">• Tasks A and B
11:00 – 11:15	Break
11:15 – 12:00	Round 1 Feedback & Round 2 Judgments for Multiple-Choice
12:00 – 1:00	Lunch
1:00 – 2:55	Round 1 Feedback & Round 2 Judgments for Multiple-Choice (continued)
2:55 – 3:00	Break
3:00 – 3:30	Round 1 Feedback & Round 2 Judgments for Constructed-Response
3:30 – 4:00	Specification Judgments
4:00 – 4:15	Feedback on Round 2 Recommended Cut Score
4:15 – 4:30	Complete Final Evaluation
4:30 – 4:45	Collect Materials; End of Day 2

Description of a Just Qualified Candidate
RVE: Reading Specialist (0304)
(Developed for the Virginia Department of Education)

A JQC ...

1. knows the characteristics of types of tests (e.g., diagnostic, formative, summative, screening) and understands “common” reading assessments (e.g., sight words, fluency, comprehension, running record)
2. knows how to interpret data to determine strengths and weaknesses in order to plan appropriate instruction for groups and individuals
3. understands the development of oral language and communication skills as it relates to reading and knows a variety of instructional strategies to support oral language and communication skills development for all students
4. understands the concepts of phonological awareness and its relationship to beginning reading, and knows instructional strategies to promote phonemic awareness
5. understands concepts of print and basic phonetic principles as they relate to reading development and knows a variety of instructional strategies to promote student application of concepts of print and phonetic principles
6. understands explicit, systematic phonics instruction and its direct correlation to reading development, and knows a variety of strategies to promote reading development at all ages
7. understands word analysis skills (e.g., word referents, meaning clues) and knows a variety of instructional strategies to enhance vocabulary knowledge to promote reading comprehension
8. understands the role of automatic word recognition (automaticity) and fluency and knows a variety of instructional strategies to promote fluency and comprehension
9. understands reading comprehension strategies for fiction and poetry including text structures and features, and knows a variety of instructional strategies for before, during, and after reading
10. understands reading comprehension strategies for informational text including text structures and features, and knows a variety of instructional strategies for before, during, and after reading
11. knows how to select a wide variety of fiction and nonfiction literature at identified reading instructional levels for all students
12. understands writing skills and processes and knows instructional strategies for promoting students’ writing development
13. understands the steps in the development of writing as a process
14. knows how to promote students’ knowledge of correct spelling, usage, and other writing mechanics and knows instructional strategies to promote student understanding of spelling, usage, and writing mechanics
15. understands writing and reading as tools for inquiry and research (e.g., reference materials, media) and knows instructional strategies to promote student understanding of writing and research
16. knows specialized knowledge and skills required to perform the role of a reading specialist (i.e., student assessment, remediation, resource to teachers)
17. knows the leadership role of the reading specialist in organizing and supervising reading programs and promoting staff development
18. knows strategies for communicating and collaborating with all members of the educational community to address the reading program

Figure 1

Panelists Names and Affiliations (RVE: Reading Specialist Standard Setting Panel)

Panelist

Terri Bredamus
Frieda E. Cason
Karen Fabrie
Stacey M. Goode
Kindel Holloman
William I. Jones
Gayle R. Kelley
Kathie Carwile Morgan
Dawn M. Plum
Kenneth Schmidt
Vickie K. Sessoms
Christi Stapleton
Susan Thompson
Katherine Wiesendanger
Carol Williams
Joyce Winfield-Reeves

Affiliation

Henry County Schools
Mack Benn, Jr. Elementary School (Suffolk)
Roanoke County Public Schools
Norfolk Public Schools
Norfolk Public Schools
Washington County Public Schools
Arlington Public Schools
Liberty University
Henrico County Public Schools
Magruder Elementary School (York County)
Sealston Elementary School (King George County)
Scott County Public Schools
Lynchburg College
Longwood University
Venable Elementary School (Charlottesville)
Clays Mills Elementary School (Halifax County)

Table 4**Specification Judgments (RVE: Reading Specialist Standard Setting Panel)**

	Very Important		Important		Slightly Important		Not Important		
	N	%	N	%	N	%	N	%	
I. Assessment and Diagnostic Teaching									
• Understand the characteristics and uses of assessment and screening measures for evaluating students' language proficiency and reading skills	12	75%	4	25%	0	0%	0	0%	
• Understand the use of assessment data to plan reading instruction	15	94%	1	6%	0	0%	0	0%	
II. Oral Language and Oral Communication									
• Understands the development of oral language and oral communication skills	5	31%	8	50%	3	19%	0	0%	
• Understand the development of phonological awareness, including phonemic awareness	13	81%	3	19%	0	0%	0	0%	
III. Reading Development									
• Understands how to promote students' understanding of concepts of print and basic phonetic principles	12	75%	4	25%	0	0%	0	0%	
• Understand explicit, systematic phonics instruction	9	56%	6	38%	1	6%	0	0%	
• Understand word-analysis skills and vocabulary development	10	63%	6	38%	0	0%	0	0%	
• Understands the development of reading fluency and reading comprehension	12	75%	4	25%	0	0%	0	0%	
• Understand reading comprehension strategies for fiction and poetry	10	63%	6	38%	0	0%	0	0%	
• Understand reading comprehension strategies for nonfiction	11	69%	5	31%	0	0%	0	0%	

Table 4 (continued)

Specification Judgments (RVE: Reading Specialist Standard Setting Panel)

	Very Important		Important		Slightly Important		Not Important	
	N	%	N	%	N	%	N	%
IV. Writing and Research								
• Understand writing skills and processes	8	50%	7	44%	1	6%	0	0%
• Understand how to promote students' knowledge of correct spelling, usage, and other writing mechanics	8	50%	6	38%	2	13%	0	0%
• Understand writing and reading as tools for inquiry and research	4	25%	11	69%	1	6%	0	0%
V. Specialized Knowledge and Leadership Skills								
• Understand specialized knowledge and skills required to perform the role of a reading specialist	10	63%	6	38%	0	0%	0	0%
• Understand leadership roles of the reading specialist in organizing and supervising reading programs and promoting staff development	8	50%	7	44%	1	6%	0	0%
• Understand strategies for communicating and collaborating with all members of the educational community to address the goals of the reading program	6	38%	8	50%	2	13%	0	0%

Table 4 (continued)

Specification Judgments (RVE: Reading Specialist Standard Setting Panel)

	Very Important		Important		Slightly Important		Not Important	
	N	%	N	%	N	%	N	%
VI. Analysis of Specialized Knowledge and Leadership Skills								
<ul style="list-style-type: none"> The candidate will apply knowledge of the elements of reading, reading instruction, and leadership skills to prepare an organized written response to a constructed-response question. 	5	31%	10	63%	1	6%	0	0%
VII. Integrated Knowledge and Understanding								
<ul style="list-style-type: none"> The candidate will apply knowledge of reading assessment and instruction to prepare an organized written response to a case study of an elementary school student. 	10	63%	6	38%	0	0%	0	0%

Table 5**Final Evaluation (RVE: Reading Specialist Standard Setting Panel)**

	Strongly Agree		Agree		Disagree		Strongly Disagree	
	N	Percent	N	Percent	N	Percent	N	Percent
• I understood the purpose of this study.	14	88%	2	13%	0	0%	0	0%
• The instructions and explanations provided by the facilitators were clear.	14	88%	2	13%	0	0%	0	0%
• The training in the standard setting method was adequate to give me the information I needed to complete my assignment.	15	94%	1	6%	0	0%	0	0%
• The explanation of how the recommended cut score is computed was clear.	15	94%	1	6%	0	0%	0	0%
• The opportunity for feedback and discussion between rounds was helpful.	13	81%	3	19%	0	0%	0	0%
• The process of making the standard setting judgments was easy to follow.	10	63%	6	38%	0	0%	0	0%

Table 5 (continued)

Final Evaluation (RVE: Reading Specialist Standard Setting Panel)

How influential was each of the following factors in guiding your standard setting judgments?	Very Influential		Somewhat Influential		Not Influential			
	N	Percent	N	Percent	N	Percent		
• The definition of the JQC	14	88%	2	13%	0	0%		
• The between-round discussions	8	50%	8	50%	0	0%		
• The knowledge/skills required to answer each test question	14	88%	2	13%	0	0%		
• The cut scores of other panel members	2	13%	10	63%	4	25%		
• My own professional experience	13	81%	2	13%	1	6%		
	Very Comfortable		Somewhat Comfortable		Somewhat Uncomfortable		Very Uncomfortable	
	N	Percent	N	Percent	N	Percent	N	Percent
• Overall, how comfortable are you with the panel's recommended cut score?	10	63%	6	38%	0	0%	0	0%
	Too Low		About Right		Too High			
	N	Percent	N	Percent	N	Percent		
• Overall, the recommended cut score is:	3	19%	13	81%	0	0%		