

Final Report on the 2016 SAT Multi-State Standard Setting

Deanna L. Morgan
Kevin Sweeney
Rosemary Reshetar
Priyank Patel
Janeen McCullough

Executive Summary

In June 2016, College Board Research and Development facilitated a Multi-State Standard Setting for the Math and Evidence-Based Reading and Writing Sections of the SAT. The purpose was to obtain three recommended cut scores for each of these exam sections that can be used for accountability purposes. This report summarizes the procedures used to obtain recommended cut scores from the standard setting panels, as well as the final cut scores that were agreed upon by the four states: Connecticut, Delaware, Maine, and New Hampshire. First, an overview of the standard setting meeting is presented, followed by a detailed description of the procedures and results.

A standards verification meeting was held separately for the SAT Essay to meet the needs of Delaware and Maine. The standards verification is described in Appendix J.

Final Report on the 2016 SAT Multi-State Standard Setting

In June 2016, College Board Psychometric and Assessment Design and Development staff conducted panel-based standard setting meetings for the new SAT. The meetings were held concurrently on June 15th – 16th, 2016 in Windsor Locks, CT at the Sheraton Hartford Hotel at Bradley Airport. The purpose of the standard setting meeting was to produce recommended cut scores on the SAT Math section and the SAT Evidence-Based Reading and Writing section (ERW) for classifying students into 4 performance levels to be used by Maine, New Hampshire, Connecticut, and Delaware for accountability purposes. This report summarizes the procedures used to collect recommended cut scores from the standard setting panelists, along with the results from the meetings. First, the instrument and participants are described. Then, procedures used during the standard setting meeting are presented, followed by a description of the results.

The New SAT

The new SAT consists of 3 sections: Math, ERW, and an optional essay. The essay was not part of the process being described in this report. The Math section has two parts: a No-Calculator part with 20 items where examinees are allotted 25 minutes and a Calculator part with 38 items and 55 minutes allotted. Both parts contain 4-option multiple choice (MC) items and student produced response (SPR) items. The ERW section also has two parts: a Reading part with 52 MC items where examinees are allotted 65 minutes, and a Writing and Language part with 44 MC items and 35 minutes. The Math and ERW sections are each on the 200-800 scale score metric. The new SAT is rights-scored meaning there is not penalty or correction for incorrect answers or guessing.

Subject Matter Experts

Two standard setting panels of subject matter experts (SMEs) were convened, one for Math and one for ERW. Each state was responsible for recruiting participants from their state in each content area. The primary requirements for participation were teaching experience of at least 3 years and content expertise in the subject area. In addition, states were provided guidance in terms of other characteristics that should be taken into consideration to obtain a panel with as much diversity as possible, to bring multiple perspectives to the meeting, and to aid in the generalizability of the results. The panels convened for Math and ERW contained 14 and 17 SMEs, respectively. At the beginning of the meetings, following a brief introduction, the SMEs were asked to complete a biographical data form for use in summarizing panelist characteristics as evidence of procedural validity (Kane, 2001; Pitoniak and Morgan, 2012). This self-report biographical data are summarized in Tables 1 and 2. In this report, the term SMEs is used interchangeably with *panelists*.

Standard Setting Meeting Procedures

In this section, the procedures used to collect standard setting ratings from the SMEs are described. The sequence of activities in this report matches that used in the June

standard setting meeting. The agenda used to guide the meeting is presented in Appendix A. The agenda for only one meeting is provided but the activities were the same for both Math and ERW.

Introduction to the Standard Setting Meeting

At the start of the meeting, panelists completed a short biographical data form (see Appendix B), for the purpose of contributing to the documentation of the procedural validity (Kane, 2001; Hambleton, Pitoniak, & Copella, 2012; Pitoniak and Morgan, 2012) of the standard setting process. Findings are summarized in Tables 1 and 2. Additionally, panelists were required to sign a confidentiality form, though the test form was a released form it was important to keep other information presented and results confidential since they were not final or released publicly yet.

At the start of the meeting, the facilitators, Kevin Sweeney for Math and Rosemary Reshetar for ERW, provided an introduction to the concept of *cut scores*, which are values used to classify student exam performance into distinct categories. In the context of the SAT Multi-State Meetings, three cut scores were used to assign examinees to one of 4 performance levels where Level 1 was low and Level 4 was high (See Figure 1).

Prior to the meeting, electronic copies of the Agenda and the Achievement Level Descriptors (ALDs) were sent to panelists, who were encouraged to review the materials and identify questions to discuss prior to completing the standard setting task. After a brief introduction of the purpose of the meeting, panelists were given an overview of the exam format by College Board Assessment Design and Development staff who remained in the room during the meeting to address any content related questions that were raised. Chris Lazarro and Laura Smith presented to the ERW panel and Andrew Schwartz and Ilirjan Cane presented to the Math panel.

Experiencing the Exam

In order to provide panelists with a frame of reference for considering student performance, the panelists took the relevant sections of the SAT in a timeframe that was reduced from that allowed operationally. Panelists did not have access to answer keys during the exam administration. Operationally, students are allotted approximately 100 minutes to complete the ERW section with 65 minutes devoted to the Reading items and 35 minutes devoted to the Writing and Language items; standard setting panelists took the ERW section in 90 minutes. Operationally, students are allotted approximately 80 minutes for the Math section with 25 minutes for the No Calculator items and 55 minutes for the Calculator items; standard setting panelists took the Math section in 70 minutes. This activity was designed to familiarize panelists with the exam questions, as well as, the rigor and time constraints experienced by students who take the exam.

Following completion of the exam, an answer key was distributed to panelists so they could grade themselves on the items. No record of SME performance on the exam was

kept, and panelists were free to share their performance with the other members of the panel at their own discretion.

Review of items. Panelists then had an opportunity to review and discuss items that they found especially difficult or confusing with an emphasis on characteristics of specific items. Panelists were reminded that the purpose of this discussion was to discuss their perceived difficulty of items in the context of the entire experience, rather than an opportunity to critique the items or the test. Any comments of a critical nature or editorial type beyond the scope of the standard setting task were to be collected and shared with College Board Assessment Design and Development staff. In general, the group of standard setting panelists was positive about the exam.

Achievement Level Descriptors

Following the exam experience and discussion, Assessment Design and Development staff reviewed the Achievement Level Descriptors (ALDs) written in collaboration with the content experts from the 4 states. ALDs describe the borderline knowledge, skills, and abilities that are required for a student to be placed into each performance level. Discussion of ALDs prior to assigning standard setting ratings helps to establish a common understanding across standard setting panelists of the meaning of the borderline of each performance level in terms of what students at the borderline know and are able to do. In essence, the ALDs serve as anchors during the rating task. The ALDs used for the standard setting are provided in Appendix C.

Borderline Examinee. Borderline examinees are students whose knowledge, skills, and abilities represent the minimal level of competence required for placement in each performance level. This concept is illustrated in Figure 2. The ALDs provided to the panelists were written with the Borderline Examinee in mind. The concept of the Borderline Examinee was reviewed thoroughly with the standard setting panelists because understanding this concept is integral to the standard setting task.

Rating Task for Standard Setting

A variety of methods have been proposed for setting performance standards on educational assessments. Despite procedural similarity across many standard setting techniques (Hambleton, Pitoniak, & Copella, 2012), Cizek (2012) describes at least ten separate standard setting processes with a host of modifications that yield even more methods that can be used to collect ratings from panelists. In spite of the numerous methods, various modifications described as Angoff standard setting procedures remain among the most widely used (Angoff, 1971; Plake & Cizek, 2012). It should be noted that the Angoff methods derive from a brief description and footnote in the 2nd Edition of Educational Measurement and is typically not implemented as originally described, thus most of the methods are more accurately referenced as Modified Angoff methods. The Angoff method and its variations are criterion-referenced standard setting methods that require panelists to estimate the probability that a “minimally acceptable person” (i.e., a borderline examinee) will answer an item correctly. These probabilities are then summed

to produce recommended cut scores. A Modified Angoff standard setting method (Plake & Cizek, 2012) was used to collect panelist ratings for the SAT Multi-State Standard Setting.

Training on the Modified Angoff Method. Panelists were trained to provide Modified Angoff ratings. The training was facilitated by the standard setting facilitator, and a variety of activities was used in order to train panelists, evaluate understanding, and provide feedback prior to the collection of actual standard setting ratings. Training materials were presented orally and visually, using PowerPoint displays and handouts.

First, the concept of a borderline examinee was reviewed. Using ALDs to represent borderline examinees in each performance level (1 to 4), panelists were asked to provide expected probabilities for correctly answering each item. In order to ease the cognitive demand during rating, panelists were asked to imagine a group of 100 borderline students at the threshold of each performance level, and estimate the number who could correctly answer each item given the knowledge, skills, and abilities required for a correct response and what the ALDs provided indicate about the knowledge, skills, and abilities of the borderline examinee at that level. Students in borderline groups were described in terms of “cuts” that distinguish between performance levels. These groups were described as follows:

- **Examinees at the 1/2 cut:** borderline examinees for Level 2
- **Examinees at the 2/3 cut:** borderline examinees for Level 3
- **Examinees at the 3/4 cut:** borderline examinees for Level 4

Panelists were restricted to ratings between 20 and 95 in intervals of 5. They were not allowed to provide ratings below 20 in order to prevent a cut score that would allow for a student to place into a performance level above Level 1 by random guessing. Similarly, panelists were not allowed to provide ratings greater than 95 in recognition that perfect performance is not common, nor a reasonable expectation of the borderline examinees. Additionally, this helps to control for examinees being required to earn a perfect score to be placed into the highest score category. While the above information is accurate for the ERW panelist ratings, it was necessary to vary the restriction on ratings below 20 for the Math panel. The Math panel was allowed to provide ratings down to zero because a portion of the test items are student produced responses (SPRs) which require examinees to write a response rather than select from a list of options, as a result, chance is not a factor in the performance on these items. However, it was later decided to allow panelists ratings to go to zero on all Math items and not just the SPRs since inspection of the percent correct values provided after Round 1 of ratings indicated extremely low performance on some items making it a realistic expectation that the borderline examinee at the lower cut scores may be expected to do even worse than chance.

Because three cut scores were needed to assign students to the 4 performance levels, panelists provided ratings for the three borderline groups simultaneously on each item. Specifically, panelists examined each item, and provided ratings for each of the three borderline groups using the following directions to guide the judgmental process:

*For each item indicate the number of borderline examinees out of 100 at each cut score that would answer the item correctly. Use only numbers between **20** and **95** in **5 point** intervals. The following list provides the possible values that can be used:*

20	25	30	35	40	45	50	55
60	65	70	75	80	85	90	95

Emphasis was placed on providing estimates of what students in each group *would* do on an exam, rather than using the ratings to describe their opinions of what students *should* do.

Next, instructions were provided for recording ratings. In order to facilitate data analyses and the provision of feedback between rounds of ratings, panelists entered ratings on individual laptops using customized Google Docs spreadsheets. Panelists were also asked to also complete a paper rating form for documentation, to serve as backup in case of technology failure, and to ease the ability to move to a new room location but still have access to their ratings for the small group discussions that would occur following Round 1. An example of a Google Docs Excel file format is given in Appendix D. The Excel worksheet included a row for each item, with three columns used to enter ratings for each borderline examinee groups. Cells for ratings were restricted to the appropriate values for each item (20-95 in intervals of 5 for MC). Panelists could either enter ratings manually, or select a value from a pull-down menu. Panelists were encouraged to use the entire range of values for the items. However, the worksheet was programmed to warn panelists of chance ratings on MC items (20 or 25) using a yellow-colored cell. Rating cells were also programmed to prevent decreasing ratings across categories. In other words, ratings were required to either increase or remain the same as borderline examinee groups increased. As a result, panelists were prevented from providing cut scores that allowed higher performance levels to be associated with lower performance.

Instructions for completing the rating form were given and the SMEs were asked to practice the method using 4 items from the SAT exam they experienced earlier. The 4 items were selected to represent different types of items (e.g., MC or SPR) and different content or testing conditions (e.g., Calculator and No Calculator, Reading and Writing). When the training round of ratings was complete, selected panelists were asked to volunteer their ratings. Then a brief discussion was held to discuss any discrepancies in ratings with emphasis in relating the discussion to the ALDs. When several panelists expressed concern that the group of panelists did not agree on all of the ratings, they were reminded that the standard setting panel was designed to represent a variety of perspectives that are needed to inform decisions about cut scores and that consensus is not a goal.

Following this discussion, panelists were asked to complete the Training Evaluation Form (Appendix E) which required them to indicate whether they were ready to proceed to the operational task. Ratings on the evaluation form were reviewed and all SMEs indicated that they understood the task and were ready to proceed. After receiving confirmation that the panelists were ready to proceed, the operational standard setting task began.

Ratings and Discussion

Next, three rounds of ratings were collected. Discussion was facilitated and feedback was provided between rounds. A description of the procedures used to collect ratings is provided below, followed by a summary of the results from the three rounds of ratings. Results from panelist evaluations of the procedures are provided in the next section.

Round 1. Round 1 ratings were provided for the entire set of items in the same fashion as was used to assign the practice ratings. When panelists were completing their ratings, Google Docs saved data every few seconds allowing the progress to be monitored in real time by the data specialist. At the end of Round 1 ratings, panelists completed a Round 1 Evaluation Form (See Appendix F) to describe their level of understanding and comfort with the standard setting task, and to identify any areas where additional discussion or training may be necessary prior to Round 2.

Feedback and Discussion for Round 1. Files from all panelists were saved in real time allowing for immediate analysis when the last panelists finished rating. Using prepared SAS code, the ratings were summarized at the individual and group level for each item and across items to provide feedback data for Round 1. Appendix H displays an example of the individual feedback forms provided to each panelist, and Appendix I shows the item-level feedback form, which includes median ratings from the entire panelist group along with observed difficulty values (p -values) from the operational administration for each item.

Discussion following Round 1 judgments utilized the observed p -value for each MC item to describe item difficulty for the exam-taking population in March 2016(see Appendix I). Items were chosen for discussion based on those which were most discrepant across panelists, as well as any additional items that panelists wished to discuss.

Next, panelists were split into small groups for discussion using a color coding system prepared in advance to ensure the groups were a mix from each of the 4 states. In the small groups, panelists were encouraged to select those items with the most discrepant ratings within the small group and in comparison to the total group for discussion, which was to focus on the rationale used to assign ratings. As appropriate, SMEs were encouraged to discuss the observed item difficulty and individual item characteristics, such as format and content.

Following the small-group discussions, panelists came together for a large-group discussion to allow the small groups to hear the rationales of the groups about items where the discrepancies were largest. Panelists were given the opportunity to suggest for the large group discussion any item(s) that they wanted to discuss or found particularly difficult to make a rating upon. In addition, the facilitator chose several items with discrepant ratings for discussion. When sufficient discussion had occurred, panelists began Round 2 of ratings. No impact data on student performance in relation to the Round 1 recommended cut scores was provided to panelists during the Round 1 discussion and feedback.

Round 2. During Round 2, panelists were instructed to review each item to confirm their rating provided in Round 1 or to provide new ratings as they deemed appropriate based on the information that was presented during the discussion. Panelists were asked to change any ratings they desired from those assigned during Round 1 in their Google Docs Spreadsheet for Round 2. Data from Round 1 of ratings were saved to a master file for each panelist before Round 2 ratings began so that the results from Round 1 were available for documentation.

Feedback and Discussion for Round 2. Feedback for Round 2 was presented to the panelists in a large-group discussion, and included the average and median cut scores from the entire group of panelists, along with measures of variability within the group. In addition, impact data were presented that described the distribution of students within each performance level, based on the recommended cut scores from Round 2, and presented the percentage of students above and below the SAT benchmark as a point of comparison. Impact data provided to the panelists was based on the aggregate of the data for the 4 states. Results for individual states were not shared with the panelists. An additional piece of information provided to panelists at this point was the percentage of students in the aggregate group at or above the SAT Benchmark. The SAT Benchmark information was provided only as a point of comparison. SAT Benchmark data is provided in Table 15. The discussion proceeded in a similar fashion to the large-group discussion following Round 1 – with an emphasis on individual items and student impact. Following the Round 2 discussion, a third and final round of ratings was collected and the results from Round 3 served as the recommendations provided to the states.

Following Round 3, panelists were asked to complete a final evaluation form (see Appendix G) to provide additional evidence of the procedural validity of the standard setting meeting.

Evaluating the Standard Setting Procedures

Panelists were asked to complete a rating form after Round 1 and a final evaluation form at the end of the standard setting meeting. Evaluation forms are shown in Appendices E-G, and results are summarized in Tables 3 and 4 for Round 1 and Tables 5 – 12 for the final evaluation.

Round 1 Evaluation

The Round 1 Evaluation Form (Appendix F) was completed by panelists immediately following Round 1 ratings and prior to any discussion of the ratings or results. The primary purpose of this evaluation was to gather evidence about panelist confidence and comfort with the rating task. In addition, panelists were asked to indicate what factors they were finding influential in making their ratings and provided an opportunity to ask for additional information about the process. The completed Round 1 Evaluation Forms were examined prior to the Round 1 discussion in order to determine if any retraining was needed.

Findings from the Round 1 Evaluation are provided in Tables 3 and 4. The first set of evaluation questions asked panelists to respond on a scale from *Strongly Disagree* (1) to *Strongly Agree* (4). The mean ratings for these 6 questions were generally high, and ranged from 3.1 to 3.8 for Math and 2.9 to 3.1 for ERW. The second set of Round 1 evaluation questions asked panelists to describe influences on their Round 1 ratings using a scale from *Not Influential* (1) to *Very Influential* (3). Relatively high average ratings at or above 2.4 were received in this section though a lower mean rating was observed related to the influence of test consequences for students (#10; Mean =1.6 for Math and 2.1 for ERW).

Panelists were also asked to provide comments about additional factors that they considered when making the Round 1 ratings. However, no comments were provided.

Final Evaluation

Following Round 3, panelists completed a Final Evaluation Form, which was comprised of 4 sections. Summaries of findings from the final evaluation (see Appendix G) are presented in Tables 5 - 12.

Section 1. First, panelists were asked to indicate the extent to which they agree or disagree with a series of statements about the standard setting process. Findings for Section 1 are summarized in Tables 5 and 6. The rating scale ranged from *Strongly Disagree* (1) to *Strongly Agree* (4) with mean ratings from 2.6 to 43.7 for Math and 2.9 to 3.8 for ERW. In addition to responding to the evaluation form questions, panelists were asked to provide comments about why they assigned ratings of *Disagree* or *Strongly Disagree*. These comments are listed at the end of each table and focus on many topics but a primary focus seems to be the rigor and use of the ALDs and the provision of impact data.

Section 2. Section 2 of the Final Evaluation Form is summarized in Tables 7 and 8. This set of questions asked panelists to indicate how influential they found specific activities or pieces of information using a rating scale that ranged from *Not Influential* (1) to *Very Influential* (3). Mean ratings ranged from 1.9 to 2.9 for Math and 2.2 to 2.9 for ERW indicating the majority of examinees rated the tasks/materials as Influential or Very Influential. Taking the test (#16; Mean = 2.9) received the highest ratings on both Math and ERW and impact data (#18; Mean = 1.9 for Math and 2.2 for ERW) received the lowest, though it was still in the Influential to Very Influential range. Lower mean ratings were also seen in Math for the average ratings of other panelists (#22, Mean=1.9). Panelists were asked to provide feedback about additional influential factors; responses are listed at the end of each table and were varied with no consistent primary focus.

Section 3. Section 3 asks panelists about the usefulness of specific activities or pieces of information that were provided with a rating scale that ranges from *Not Useful* (1) to *Very Useful* (3). Findings are summarized in Tables 9 and 10. Mean ratings of usefulness ranged from 2.1 to 2.9 for Math and 2.5 to 3.0 for ERW, with taking the exam (#27; Mean = 2.9), receiving the highest ratings for Math and item difficulty values (#33; Mean = 3.0) receiving the highest ratings for ERW. An opportunity to provide feedback on other

information that would have been useful to the panelists was given on the evaluation form and those comments are summarized at the end of each table.

Section 4. The final section of the Final Evaluation Form asked panelists to describe the amount of time devoted to each activity using a scale ranging from *Too Little Time* (1) to *Too Much Time* (3). Findings are summarized in Tables 11 and 12. Mean ratings ranged from 1.9 to 2.3 for Math and 1.8 to 2.2 for ERW. All mean ratings centered around 2.0 indicating that the timing was generally about right for most panelists. Finally, panelists were asked to provide additional comments about the standard setting procedure; these comments are provided at the end of each table.

Setting the Final Cuts

Ratings from all 3 rounds were provided to the states, along with the Standard Error of Judgment (SEJ) from each as an indicator of the variability in panelist ratings. The Standard Error of Measurement (SEM), 30.00, was also provided as another measure of variation. The SEJ should get smaller across rounds as the agreement among panelists increases. The SEJ and the SEM may be used to make adjustments to the recommended cut scores when there is a clear rationale for doing so. Generally, adjustments to the cut scores should stay within +/- 1.5 SEJ or SEM to maintain the integrity of the panel.

Table 13 provides a summary of the cut scores (mean, median, min, and max) based on the three rounds of standard setting ratings. The SEJ for ratings during each round are provided in Table 14. SEJ values are on the SAT scale score metric which proceeds by multiples of 10. The majority of the SEJ values are below 10 indicating they are relatively small and there is general agreement among panelists. The SEJ for Math Level 2 in Round 3 has a slight increase but remains very small. The SEJ for ERW Level 3 increases slightly in Round 3 but remains very small, however, the SEJ for ERW Level 4 increases in each round indicating panelists are becoming more discrepant. This trend is consistent with anecdotal feedback from observers that panelists in ERW were having difficulty coming to consensus about expectations for students and in particular reconciling the different panelists perspectives on expectations with the consequences for students as presented by the impact data.

Table 16 summarizes the expected student impact based on the recommended cut scores from Round 3. Finally, Table 17 provides the final cut scores agreed upon by the four states and the expected impact in the aggregate and by state.

Concluding Comments

Overall, the standard setting meeting went very well. Procedurally, everything went as scheduled and expected. The cut scores recommended by the panel during the third round of ratings were provided to the state departments of education staff for consideration along with other information they considered relevant. The four states met

numerous times and came to final agreement on the set of cut scores as listed in Table 17. In a separate meeting, panelists from Delaware and Maine convened for a standards verification meeting for the SAT Essay. That process and the results are described in Appendix J.

References

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement*, (2nd ed., pp. 508-560), Washington, DC: American Council on Education.
- Cizek, G.J. (2012). *Setting performance standards: Foundations, Methods, and Innovations*. New York: Routledge
- Hambleton, R.K., Pitoniak, M.J., & Copella, J.M. (2012). Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of results. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, Methods, and Innovations*, (pp. 47-76). New York: Routledge.
- Kane, M. (2001). So much remains the same: Conceptions and status of validation in setting standards. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*, (pp. 53-88). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pitoniak, M. J., & Morgan, D.L. (2012). Setting and validating cutscores for tests. In C. Secolsky & D.B. Denison (Eds.), *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (pp.343-366). New York, NY: Routledge.
- Plake, B.S., & Cizek, G.J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no standard setting methods. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*,(pp. 181-200). Mahwah, NJ: Lawrence Erlbaum Associates.

Table 1*Biographical Information for SAT Math Standard Setting Panel*

Survey Topic	Response Categories	N (Total N = 16#)	%
Demographic Characteristics			
Gender	Male	5	31.25
	Female	11	68.75
Race/ethnicity	American Indian or Alaska Native	0	0
	Asian, Asian American, or Pacific Islander	0	0
	Black or African American	1	6.25
	Mexican or Mexican American	0	0
	Puerto Rican	0	0
	Other Hispanic, Latino, or Latin American	0	0
	White	15	93.75
	Other	0	0
Education	Undergraduate Degree	1	6.25
	Master's Degree	11	68.75
	Specialist's Degree	1	6.25
	Doctoral Degree	2	12.50
	Other Education Level	1	6.25
Teaching Experience			
Years of Experience Teaching Chemistry	1-3 years	0	0
	4-6 years	1	6.25
	7-12 years	5	31.25
	12+ years	10	62.5
Current Teaching Level	High School	10	62.5
	College	0	0
	Both College & High School	6	37.50
	Other	0	0
Special Interest Groups*			
	Special Education	2	12.50
	English Language Learners	0	0
	Gifted and talented	0	0
	Community member	0	0
	Postsecondary	0	0
	Other	1	6.25

*Categories are not necessarily mutually exclusive.

#One panelist did not turn in their form.

Table 2

Biographical Information for SAT Evidence-Based Reading and Writing Standard Setting Panel

Survey Topic	Response Categories	N (Total N = 14)	%
Demographic Characteristics			
Gender	Male	4	28.57
	Female	10	71.43
Race/ethnicity	American Indian or Alaska Native	1	7.14
	Asian, Asian American, or Pacific Islander	1	7.14
	Black or African American	2	14.29
	Mexican or Mexican American	0	0
	Puerto Rican	0	0
	Other Hispanic, Latino, or Latin American	0	0
	White	10	71.43
	Other	0	0
Education	Undergraduate Degree	1	7.14
	Master's Degree	9	64.29
	Specialist's Degree	1	7.14
	Doctoral Degree	3	21.43
	Other Education Level	0	0
Teaching Experience			
Years of Experience Teaching Chemistry	1-3 years	1	7.14
	4-6 years	2	14.29
	7-12 years	2	14.29
	12+ years	9	64.29
Current Teaching Level	High School	8	57.14
	College	1	7.14
	Both College & High School	5	35.71
	Other	0	0
Special Interest Groups*			
	Special Education	1	7.14
	English Language Learners	0	0
	Gifted and talented	1	7.14
	Community member	0	0
	Postsecondary	3	21.43
	Other	1	7.14

*Categories are not necessarily mutually exclusive.

Table 3.
Summary of Round 1 Evaluation – Math

Rating Scale	Questions		Mean Rating
Strongly Disagree (1) to Strongly Agree (4)	1	I understand the purpose of the study.	3.8
	2	The instructions and explanations provided by the facilitator were clear.	3.1
	3	The training in the standard setting method gave me the information I needed to complete my assignment.	3.4
	4	The ALDs that were developed prior to the meeting were accurate.	3.4
	5	I understand the concept of the borderline examinee.	3.4
	6	The ALDs helped me to determine how to rate each item.	3.5
Not Influential (1) to Very Influential (3)	7	Completing the test before beginning the task.	2.8
	8	My perception of the difficulty of the items	2.4
	9	The ALDs	2.5
	10	The consequences of the test for students	1.6
	11	My experience with students in my classroom	2.4

Table 4.*Summary of Round 1 Evaluation – Evidence-Based Reading and Writing*

Rating Scale	Questions		Mean Rating
Strongly Disagree (1) to Strongly Agree (4)	1	I understand the purpose of the study.	3.1
	2	The instructions and explanations provided by the facilitator were clear.	3.1
	3	The training in the standard setting method gave me the information I needed to complete my assignment.	3.1
	4	The ALDs that were developed prior to the meeting were accurate.	2.9
	5	I understand the concept of the borderline examinee.	3.1
	6	The ALDs helped me to determine how to rate each item.	3.1
Not Influential (1) to Very Influential (3)	7	Completing the test before beginning the task.	3.0
	8	My perception of the difficulty of the items	2.7
	9	The ALDs	2.6
	10	The consequences of the test for students	2.1
	11	My experience with students in my classroom	2.8

Table 5
Summary of Final Evaluation – Section 1 - Math

Rating Scale	Questions		Mean Rating
Strongly Disagree (1) to Strongly Agree (4)	1	I understood the purpose of the study.	3.6
	2	The instructions and explanations provided by the facilitator were clear.	3.4
	3	The training on the standard setting method gave me the information I needed to complete my assignment.	3.4
	4	The ALDs that were developed prior to the meeting were accurate.	3.2
	5	I understood the concept of the borderline examinee.	3.4
	6	The ALDs helped me to determine how to rate each item.	3.1
	7	It was beneficial to have an opportunity for discussion and to review feedback between rounds.	3.7
	8	The impact data showing the percent of students expected to place into each category made a difference in how I rated the items in the next round.	2.9
	9	The opportunity to more than 1 round of ratings (i.e., round 2) helped me to feel more confident about my final ratings.	3.6
	10	I felt engaged in the process.	3.6
	11	I was comfortable sharing my ideas with the other panelists during the discussions.	3.7
	12	I am confident this standard setting process will produce fair cut scores.	2.6
	13	I would be comfortable defending this process to my peers.	2.9
Extended Response	14	<i>Comments if Disagree or Strongly Disagree</i>	

Responses were as follows:

- I have concerns @ not this level in the process but the next levels. I have participated in this type panel in past & always found it a valuable experience. This one was as well.
- I found that the impact data had little influence on my round 3 scores. As a large group we discussed & came to a consensus about a number of items that were discrepant. This data created a discussion more related to cut off scores than the individual items.
- I would feel comfortable with what I did but I feel some people were thinking only what a higher level student could do which raised the cut scores a lot. They didn't understand a boardline student.
- ALDs might have been too difficult
- I believe the ALDs were created to set a high benchmark for statewide testing. I believe that combining 4 states in one room created varied responses and a wide range of student outcomes.
- The ALDs were one component, but not the most important.
- I think norms should be set up at the start of the meeting/conference. I think the cut scores maybe too high.
- The impact data on round 2, being presented after Round 2 discussion, did not really give opportunity to incorporate into final rating.
- I feel the people often considered their own students over the ALDs. Also, I think the SAT is designed for college bound

students and we are holding all students to this high level.

Table 6

Summary of Final Evaluation – Section 1 – Evidence-Based Reading and Writing

Rating Scale	Questions	Mean Rating
Strongly Disagree (1) to Strongly Agree (4)	1 I understood the purpose of the study.	3.8
	2 The instructions and explanations provided by the facilitator were clear.	3.7
	3 The training on the standard setting method gave me the information I needed to complete my assignment.	3.5
	4 The ALDs that were developed prior to the meeting were accurate.	3.1
	5 I understood the concept of the borderline examinee.	3.6
	6 The ALDs helped me to determine how to rate each item.	3.5
	7 It was beneficial to have an opportunity for discussion and to review feedback between rounds.	3.8
	8 The impact data showing the percent of students expected to place into each category made a difference in how I rated the items in the next round.	3.2
	9 The opportunity to more than 1 round of ratings (i.e., round 2) helped me to feel more confident about my final ratings.	3.7
	10 I felt engaged in the process.	3.6
	11 I was comfortable sharing my ideas with the other panelists during the discussions.	3.6
	12 I am confident this standard setting process will produce fair cut scores.	2.9
	13 I would be comfortable defending this process to my peers.	3.0
Extended Response	14	Comments if <i>Disagree</i> or <i>Strongly Disagree</i>
<p>Responses were as follows:</p> <ul style="list-style-type: none"> • The cut scores we reached are unrealistically high. Looking at impact data and difficulty level earlier/more closely might have helped. • The ALDs at the 1-2 level are problematic. • Though the impact data was telling, I still think the process called on us to focus on the ALDs and questions. • CT teachers would have been part of ALD setting. • Please prevent people from so personalizing it becomes about them 		

Table 7*Summary of Final Evaluation – Section 2 - Math*

Rating Scale	Questions		Mean Rating
Not Influential (1) to Very Influential (3)	15	Completing the test before beginning the task.	2.9
	16	My perception of the difficulty of the items	2.4
	17	The actual item difficulty provided for each item	2.1
	18	Distributions of students expected to earn each achievement level	1.9
	19	Small group discussion from after Round 1	2.7
	20	Large group discussion from after Round 1	2.2
	21	Discussion after Round 2	2.4
	22	The average ratings of other panelists	1.9
	23	The ALDs	2.3
	24	The consequences of the exam for students	2.1
	25	My experience with students in my classroom	2.4
Extended Response	26	Other factors that influenced decisions:	
Responses were as follows: <ul style="list-style-type: none"> • The consequences of the exam scores for the teacher/school. • I think we should have seen & discussed Rnd 3 results. • I don't believe supervisors should be panelist, they have a disconnect of what kids really can do. • Percentages of students scoring in each level helped fire conversations • Testing practices & Constraints all influenced my decisions. • Being able to ask people who had generated questions some insight to the "spirit" of the questions. • I feel that we should have only looked at our state data. And had a panel only of our state. 			

Table 8*Summary of Final Evaluation – Section 2 – Evidence-Based Reading and Writing*

Rating Scale	Questions		Mean Rating
Not Influential (1) to Very Influential (3)	15	Completing the test before beginning the task.	2.9
	16	My perception of the difficulty of the items	2.7
	17	The actual item difficulty provided for each item	2.8
	18	Distributions of students expected to earn each achievement level	2.2
	19	Small group discussion from after Round 1	2.6
	20	Large group discussion from after Round 1	2.5
	21	Discussion after Round 2	2.2
	22	The average ratings of other panelists	2.4
	23	The ALDs	2.7
	24	The consequences of the exam for students	2.3
	25	My experience with students in my classroom	2.5
Extended Response	26	Other factors that influenced decisions:	
Responses were as follows: <ul style="list-style-type: none"> • My familiarity and work w/ the CCSS • Test logistics, formatting, and word choice influenced my ratings. • Need to stay on topic purpose of our participation was not to debate exam's merits. ALDs can be better formatted-horizontal alignment 			

Table 9*Summary of Final Evaluation – Section 3 - Math*

Rating Scale	Questions		Mean Rating
Not Useful (1) to Very Useful (3)	27	Taking the exam prior to beginning the task	2.9
	28	Practicing the procedure with real items prior to beginning the actual rating task	2.8
	29	Referencing the ALDs	2.6
	30	Small Group Discussion after round 1	2.6
	31	Large Group Discussion after round 1	2.4
	32	Test overview	2.1
	33	Actual Item Difficulty Values	2.2
	34	Distributions of students earning each achievement level	2.2
	35	Large group discussion after round 2	2.3
Extended Response	36	Other information that would have been useful during ratings:	
Responses were as follows: <ul style="list-style-type: none"> • Pure CT scores • The SAT scaled score after Round 1 would have helped distribution of scores for some. • Small groups helped more than the large group. I wish we had jigsawed more. • The range of values and median scores to determine if there were numerous ratings high or low or one outlier. • I would have wanted to see the distribution after Round 1. 			

Table 10*Summary of Final Evaluation – Section 3 – Evidence-Based Reading and Writing*

Rating Scale	Questions		Mean Rating
Not Useful (1) to Very Useful (3)	27	Taking the exam prior to beginning the task	2.9
	28	Practicing the procedure with real items prior to beginning the actual rating task	2.9
	29	Referencing the ALDs	2.8
	30	Small Group Discussion after round 1	2.8
	31	Large Group Discussion after round 1	2.7
	32	Test overview	2.7
	33	Actual Item Difficulty Values	3.0
	34	Distributions of students earning each achievement level	2.8
	35	Large group discussion after round 2	2.5
Extended Response	36	Other information that would have been useful during ratings:	
Responses were as follows: <ul style="list-style-type: none"> Knowing test question difficulty & a cross walk document to align to ALDs tested. better explanation of percentage translation to scores/cut scores 			

Table 11
Summary of Final Evaluation – Section 4 - Math

Rating Scale	Questions		Mean Rating
Too Little Time (1) to Too Much Time (3)	37	Taking the test	1.9
	38	Reviewing the ALDs	2.0
	39	Training on the rating task for Round 1	2.2
	40	Round 1 of the rating task	2.0
	41	Small Group Discussion after round 1	2.1
	42	Large Group Discussion after round 1	2.3
	43	Round 2 of the rating task	2.0
Extended Response	44	Large group discussion after round 2	2.1
	45	Any additional comments:	
<p>Responses were as follows:</p> <ul style="list-style-type: none"> • The pacing seemed slow at times during the first morning, but after that things moved well. * more teachers on the panel. • Some panelists did not know how to do some problems. They should be taught before they rate the problems. • I think it would have been better if there was a more even distribution of the types of schools the panelists represented. Some taught in the same school! • Very interesting to learn this process. I feel like my input was valued. • I have trouble with basically "guessing" cut % of students. • I am very concerned leaving that the Round 2 % do not align to the SAT benchmark %. 			

Table 12*Summary of Final Evaluation – Section 4 – Evidence-Based Reading and Writing*

Rating Scale	Questions		Mean Rating
Too Little Time (1) to Too Much Time (3)	37	Taking the test	1.8
	38	Reviewing the ALDs	1.8
	39	Training on the rating task for Round 1	2.0
	40	Round 1 of the rating task	2.1
	41	Small Group Discussion after round 1	2.2
	42	Large Group Discussion after round 1	2.1
	43	Round 2 of the rating task	2.1
Extended Response	44	Large group discussion after round 2	2.2
	45	Any additional comments:	
Responses were as follows: <ul style="list-style-type: none"> • Complex process. Excellent job by facilitators. Thank you • Curious to see where this ends up. 			

Table 13*Cut Score Summary by Round**

Math

Round	Level 2					Level 3					Level 4				
	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD
1	440	440	390	510	29.49	570	560	520	640	40.68	690	690	650	740	29.63
2	420	430	390	460	23.08	570	560	520	610	27.57	690	690	650	720	23.01
3	420	420	380	460	23.80	550	560	510	600	26.80	680	680	650	720	21.25

*Values have been rounded to the closest reportable score

Evidence-Based Reading and Writing

Round	Level 2					Level 3					Level 4				
	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD
1	470	470	350	550	53.02	590	580	540	660	39.26	700	690	620	750	36.74
2	460	460	390	510	33.84	580	580	530	630	30.50	700	710	610	750	37.77
3	450	450	390	500	32.69	570	580	510	620	31.53	690	700	610	750	41.22

*Values have been rounded to the closest reportable score

Table 14*Standard Error of Judgment by Round (SAT Scale Score Metric)*

Math

Standard Error of Judgment	Level 2	Level 3	Level 4
Round 1	7.61	10.50	7.65
Round 2	5.77	6.89	5.75
Round 3	5.95	6.70	5.31

Evidence-Based Reading and Writing

Standard Error of Judgment	Level 2	Level 3	Level 4
Round 1	14.17	10.49	9.82
Round 2	9.04	8.15	10.10
Round 3	8.74	8.43	11.02

Table 15*SAT Benchmark Performance by 4 State Aggregate and by State.*

Frequency Percent Row Pct Col Pct	Table of MATH_BENCHMARK by ERW_BENCHMARK		
	MATH_BENCHMARK	ERW_BENCHMARK	
		0	1
0	24565 36.09 57.67 95.96	18029 26.49 42.33 42.46	42594 62.58
1	1035 1.52 4.06 4.04	24434 35.90 95.94 57.54	25469 37.42
Total	25600 37.61	42463 62.39	68063 100.00

STUDENT_STATE_CD=CT

Frequency Percent Row Pct Col Pct	Table of MATH_BENCHMARK by ERW_BENCHMARK			
	MATH_BENCHMARK	ERW_BENCHMARK		
		0	1	Total
0	12599	9874	22473	
	34.41	26.97	61.38	
	56.06	43.94		
	96.59	41.90		
1	445	13693	14138	
	1.22	37.40	38.62	
	3.15	96.85		
	3.41	58.10		
Total	13044	23567	36611	
	35.63	64.37	100.00	

STUDENT_STATE_CD=DE

Frequency Percent Row Pct Col Pct	Table of MATH_BENCHMARK by ERW_BENCHMARK			
	MATH_BENCHMARK	ERW_BENCHMARK		
		0	1	Total
0	3768	1931	5699	
	45.52	23.33	68.85	
	66.12	33.88		
	95.63	44.52		
1	172	2406	2578	
	2.08	29.07	31.15	
	6.67	93.33		
	4.37	55.48		
Total	3940	4337	8277	
	47.60	52.40	100.00	

STUDENT_STATE_CD=ME

Frequency Percent Row Pct Col Pct	Table of MATH_BENCHMARK by ERW_BENCHMARK			
	MATH_BENCHMARK	ERW_BENCHMARK		
		0	1	Total
0	4420	2982	7402	
	38.63	26.06	64.69	
	59.71	40.29		
	95.42	43.78		
1	212	3829	4041	
	1.85	33.46	35.31	
	5.25	94.75		
	4.58	56.22		
Total	4632	6811	11443	
	40.48	59.52	100.00	

STUDENT_STATE_CD=NH

Frequency Percent Row Pct Col Pct	Table of MATH_BENCHMARK by ERW_BENCHMARK			
	MATH_BENCHMARK	ERW_BENCHMARK		
		0	1	Total
0	3778	3242	7020	
	32.20	27.63	59.84	
	53.82	46.18		
	94.83	41.84		
1	206	4506	4712	
	1.76	38.41	40.16	
	4.37	95.63		
	5.17	58.16		
Total	3984	7748	11732	
	33.96	66.04	100.00	

Table 16*Impact data for Round 3 – Percent in Category Using Median Cut Scores**Math*

Group	Level 1	Level 2	Level 3	Level 4
Four State Aggregate	23.66	47.68	23.03	5.62
Connecticut	23.73	45.36	24.56	6.35
Delaware	27.20	52.01	16.78	4.01
Maine	25.74	49.77	19.97	4.53
New Hampshire	18.91	49.84	25.69	5.56

Evidence-Based Reading and Writing

Group	Level 1	Level 2	Level 3	Level 4
Four State Aggregate	28.15	43.84	24.02	3.99
Connecticut	26.57	43.30	25.68	4.45
Delaware	37.10	41.59	18.30	3.01
Maine	30.82	43.77	21.82	3.58
New Hampshire	24.19	47.20	24.99	3.61

Table 17*Final Cut Scores from Policy Meeting and Impact*

Math

Performance Level	Cut Score	Percent of Examinees in Level				
		Four State Aggregate	CT	DE	ME	NH
Level 4	650	9.57	10.57	7.02	8.23	9.58
Level 3	530	27.85	28.05	24.13	27.08	30.58
Level 2	420	38.92	37.65	41.66	38.95	40.92
Level 1	--	23.66	23.73	27.20	25.74	18.91

Evidence-Based Reading and Writing

Performance Level	Cut Score	Percent of Examinees in Level				
		Four State Aggregate	CT	DE	ME	NH
Level 4	630	14.97	16.75	10.43	13.10	14.44
Level 3	480	47.42	47.62	41.97	46.42	51.60
Level 2	420	18.38	17.63	21.29	18.75	18.31
Level 1	--	19.23	18.00	26.31	21.73	15.65

Figure 1. *Diagram of Performance Levels*

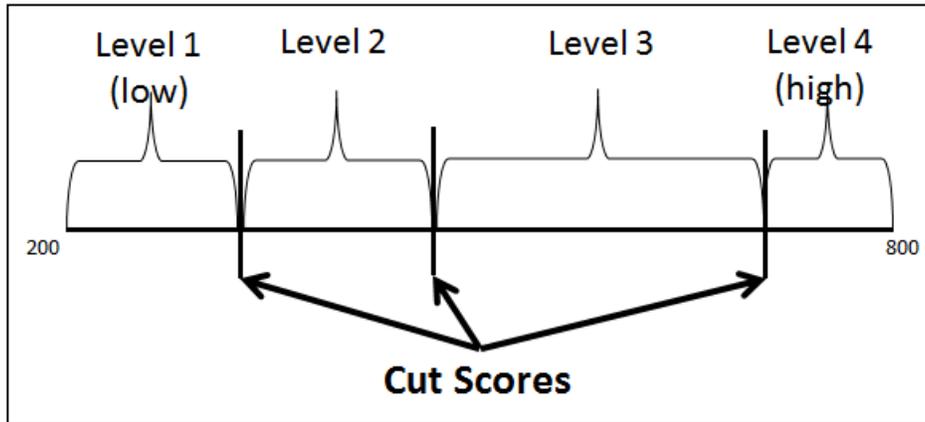
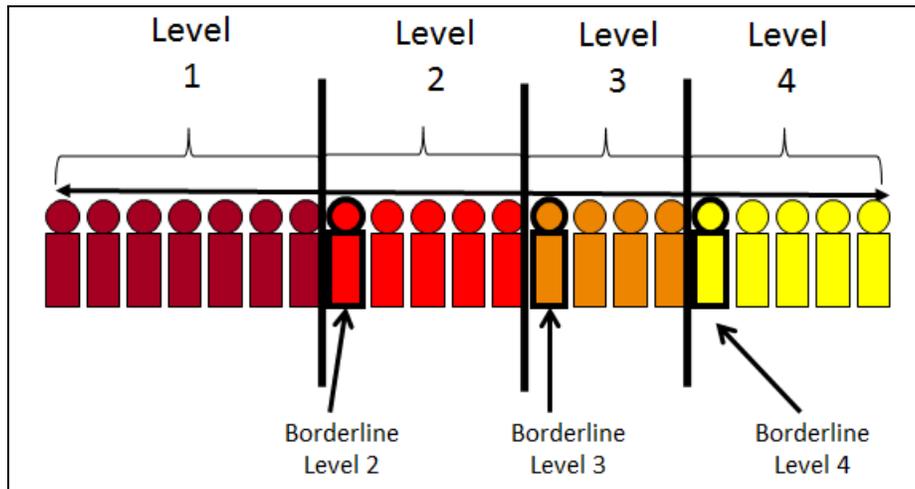


Figure 2. *Borderline Examinees*



Note. This figure depicts the concept of a *borderline examinee*, and was used during training for the standard setting described in this report. The proportions of examinees shown in each category are not intended to reflect the distribution of examinees within the four categories, actual or expected.

Appendix A: Standard Setting Meeting Agenda (Math agenda was the same)

**SAT ERW Standard Setting Agenda
June 15th - 16th, 2016**

Sheraton Hartford Hotel at Bradley Airport
1 Bradley International Airport
Windsor Locks, CT 06096

Day 1 – Wednesday, June 15th, 2016*

- 7:15 – 8:25am Breakfast is available in the Airport Concourse on the lobby level
- 8:25 – 8:30am Check In at table outside Suffield Room on the lobby level
- 8:30 – 9:30am Welcome and Introduction, Overview of SAT ERW Section and Standard Setting Process
- 9:30 – 11:00am Experience the ERW Section of the SAT
- 11:00 – 11:15am Break – Lobby Lounge
- 11:15 – 12:15pm Review of MCQ items
- 12:15 – 1:15pm Lunch – Airport Concourse
- 1:15 – 2:30pm Review of Achievement Level Descriptors (ALDs)
- 2:30 – 2:45pm Break – Lobby Lounge
- 2:45 – 3:45pm Training and Practice on Modified Angoff method
- 3:45 - 6:00pm Provide Round 1 Ratings, Evaluation Form

- 6:30 – 8:00pm Dinner at Amelia's on the Mezzanine

Day 2 – Thursday, June 16th, 2016

- 7:15 – 8:25am Breakfast is available in the Airport Concourse
- 8:25 – 8:30am Check In – Suffield

8:30 – 10:00am	Feedback and Small Group Discussion of Round 1 Ratings
10:00 – 10:15am	Break – <u>Lobby Lounge</u>
10:15 – 11:15am	Large Group Discussion
11:15 – 12:30am	Provide Round 2 Ratings
12:30 – 1:30pm	Lunch – <u>Airport Concourse</u>
1:30 – 3:00pm	Feedback and Large Group Discussion of Round 2 Ratings and Impact Data
3:00 – 3:15pm	Break – <u>Lobby Lounge</u>
3:15 – 4:30pm	Provide Round 3 Ratings
4:30 – 5:00pm	Final Evaluation Form, Check in Materials, Dismiss

Appendix B: Biographical Data Form

Biographical Data Form

Please circle the letter of the answer choice that most represents you. This information is for the reporting of panel member diversity as a measure of the generalizability and validity of the results and will be reported in aggregate form only. Data will be used for research purposes only.

1. Gender:

- A. Female
- B. Male

2. How do you describe yourself? (please choose ONE option, if more than 1 apply then please choose Multicultural)

- A. American Indian or Alaska Native
Asian, Asian American or Pacific Islander
- B. Islander
- C. Black or African American
- D. Mexican or Mexican American
- E. Puerto Rican
Other Hispanic, Latino, or Latin American
- F. American
- G. White
- H. Other
- I. Multicultural

3. Years of experience teaching Reading and/or Writing at the high school level or a corresponding course at the college level (including this year):

- A. 1 - 3 years
- B. 4 - 6 years
- C. 7 - 12 years
- D. More than 12 years

4. Levels of Classes You Teach:

- A. High School
- B. College
- C. Both College & High School
- D. Other: _____

5. Please list the name of the institution where you teach and the city, state where the institution is located:

Name of Institution: _____

City, State of Institution: _____

6. My highest level of education completed is:

- A. Undergraduate Degree
- B. Master's Degree
- C. Specialist Degree or ABD
- D. Doctoral Degree
- E. Other: _____

7. Do you represent any of the following special interest groups?

- A. Special Education
- B. English Language Learners
- C. Gifted and Talented
- D. Community Member
- E. Postsecondary
- F. Other: (Please specify) _____

Appendix C: Achievement Level Descriptors

Multi-State System of Student Assessments

ACHIEVEMENT LEVEL DESCRIPTORS

Introduction to the Multi-State Achievement Level Descriptors for Mathematics

The Mathematics portion of the Redesigned SAT assesses how well a student demonstrates college and career readiness. The **Mathematics** Test focuses on a range of skills and knowledge needed to formulate and solve problems with and without context. The Mathematics test measures a students' ability in the areas of fluency, conceptual understanding, and application. Fluency requires that students solve problems accurately, efficiently and strategically. Conceptual understanding requires students to demonstrate their understanding of math concepts, operations and relations. Finally, application involves analyzing a situation to represent and solve the problem mathematically.

The Mathematics Achievement Level Descriptors (ALDs) are aligned to the Common Core standards for high school and therefore based on the progression of the standards. The achievement level descriptors are text descriptions of the fundamental skills and knowledge demonstrated by students in each category of achievement.

ACHIEVEMENT LEVEL DESCRIPTORS

Math – GRADE 11

Achievement Level 4	Achievement Level 3	Achievement Level 2	Achievement Level 1
Advanced	Proficient/Meets Standard	Partial/Below Standard	Inadequate/Well-Below Standard
<p>The student has exceeded the achievement level and demonstrates a <i>thorough</i> understanding of, and ability to apply the mathematics knowledge and skills needed for college and career readiness and achievement relative to the Math Content Standards. The student solves problems that call for a range of strategies, accurate and insightful reasoning, and connecting different areas of mathematics.</p>	<p>The student has met the achievement level and demonstrates an <i>adequate</i> understanding of, and ability to apply the mathematics knowledge and skills needed for college and career readiness and achievement relative to the Math Content Standards. The student solves problems that call for effective use of strategies and accurate reasoning in different areas of mathematics.</p>	<p>The student has partially met the achievement level and demonstrates an <i>incomplete</i> understanding of, and ability to apply the mathematics knowledge and skills needed for college and career readiness and achievement relative to the Math Content Standards. The student solves problems that call for simple strategies and reasoning accurately applied to basic areas of mathematics.</p>	<p>The student has not met the achievement level and demonstrates a <i>minimal</i> understanding of, and ability to apply the mathematics knowledge and skills needed for college and career readiness and achievement relative to the Math Content Standards. The student solves some problems that require applying simple strategies to basic areas of mathematics without an understanding of the reasoning behind the strategies.</p>

Students at Level 1, demonstrate knowledge of simple linear equations, mostly in one-step problems in context and solve problems with given data displayed in graphs or tables. They can solve problems arising from familiar contexts and identify important quantities and begin developing models. They solve problems that involve simple logical reasoning with basic abstract concepts.

Students at that level will:

- Understand some ratio concepts and use ratio reasoning to solve problems in a variety of contexts but lacks these foundational skills to advance to more problem solving. Inconsistently, use proportional relationships to solve multistep ratio and percent problems.
- Create and solve simple linear equations in one variable but makes limited strategic use of algebraic structure, including writing equivalent equations.
- Solve a system of linear equations to solve problems in a familiar context.
- Analyze and solve pairs of simultaneous linear equations given only algebraically and graphically.
- Solve real-world and mathematical problems leading to two linear equations in two variables given only familiar contexts.
- Solve real-world and mathematical problems about a geometric figure or an object that can be modeled by a geometric figure using given information such as length, area, surface area, or volume.
- Demonstrate some procedural fluency by selecting the correct area or volume formula and correctly calculating a specified value.
- Use the relationship between variables shown mostly in a graph to make predictions and conclusions given a simple context.
- Understand and recognize linear and non-linear functions but cannot differentiate between quadratic and exponential functions.
- Interpret information from a representation of data given in a familiar context.
- Given an appropriate data set, determine the measures of center (mean, median) to summarize a data set.
- Use some concepts related to congruence and similarity of triangles to solve familiar problems.
- Apply some knowledge of the relationship of angles (complementary, supplementary, corresponding, vertical, etc...) formed when a transversal cuts parallel lines.

•
Students at Level 2, demonstrate some knowledge of more complex linear and systems of linear relationships in one and two variables with and without context. They can apply, inconsistently, their knowledge of multiple representations and the interpretation of these representations. They have some difficulty with multi-step problems that require several skills and concepts. Overall, students use a limited range of strategies needed to solve different types of problems. They can solve problems that require identifying key quantities, and recognizing and developing missing information. They can begin to identify logical assumptions within a model and produce partial justifications and explanations with the model.

In addition to the mathematical content and cognitive demand in Level 1, students will

- Apply proportional relationships, ratios, rates, and units in a wide variety of contexts. Examples include but are not limited to scale drawings and problems in the natural and social sciences.
- Solve problems involving unit conversion within measurement systems.
- Understand and use the fact that when two quantities are in a proportional relationship, if one changes by a scale factor, then the other also changes by the same scale factor.
- Use percentages to solve problems in a variety of contexts. Examples include, but are not limited to, discounts, interest, taxes, tips, and percent increases and decreases for many different quantities.
- Understand and use the relationship between percent change and growth factor (5% and 1.05, for example); include percentages greater than or equal to 100%.
- Make strategic use of algebraic structure and the properties of operations to identify and create equivalent expressions (linear, quadratic and exponential).
- Add, subtract and multiply polynomials.
- Create and use linear equations in one- and two-variables to model and solve problems in a context.
- Interpret slope (rate of change) and the intercept (constant term) of a linear model in the content of the data.
- Use the relationship between variables shown in a graph to make predictions and conclusions given a context.
- Interpret and solve a linear equation in one variable, making strategic use of algebraic structure.
- Interpret the meaning of a linear function that represents a context using two variables, explain how the quantities are related.
- Make connections between tabular, algebraic, and graphical representations of a linear equation in two variables - deriving one representation from the other.
- Write an equation for a line given different conditions (two points on the line, one point and the slope of the line, or one point and a parallel or perpendicular line).
- Create and use a system of two linear equations in two variables to solve problems in a context.
- Choose an appropriate graphical representation for a given data set.
- Summarize, represent and interpret information from a given representation of data in a variety of context.
- Given an appropriate data set, determine the measures of center (mean, median) and spread (interquartile, standard deviation) to summarize one set of data.
- Use concepts related to congruence and similarity of triangles to solve problems.
- Apply knowledge that changing by a scale factor of k changes all lengths by a factor of k , but angle measures remain unchanged.
- Apply knowledge of the relationship of angles (complementary, supplementary, corresponding, vertical, etc...) formed when a transversal cuts parallel lines.

Students at Level 3, demonstrate some mastery of linear and non-linear relationships and can apply these in different contexts. They can derive, create, and solve equations or functions to model problems in a variety of contexts with multiple variables. They can analyze data to make predictions and calculate probability concepts with and without contexts. They fluently use properties of the number system to solve equations with rational exponents and make sense of the algebraic structure to solve these problems in multiple representations. They can apply properties of right triangles to solve problems. They can construct chains of reasoning to justify a model used, produce justification of interpretations, state logical assumptions, and compare and contrast multiple plausible solutions.

OR

Overall, students can solve a range of complex, well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. They can solve problems that require them to construct, justify, and reason with mathematical models in a variety of settings. They can compare the strengths and weaknesses of different models for the same setting, and they can support their reasoning in constructing and using mathematical models.

In addition to the mathematical content and cognitive demand in Level 2, students will

- Create, use, and apply proportional relationships, ratios, rates, and units in a wide variety of contexts. Examples include but are not limited to scale drawings and problems in the natural and social sciences.
- Understand and use the fact that when two quantities are in a proportional relationship, if one changes by a scale factor, then the other also changes by the same scale factor to solve problems in different contexts.
- Make strategic use of algebraic structure and the properties of operations to identify and create equivalent expressions, (linear, rational, radicals and rational and exponential expressions)
- Add, subtract, multiply, and factor polynomials.
- Construct and use linear equations and inequalities in one or two variables to solve problems and interpret the solutions in a variety of contexts.
- Make connections between tabular, algebraic, and graphical representations of linear equations and inequalities in one or two variables by deriving one from the other.
- Solve systems of linear equations and inequalities in a variety of context.
- Make strategic use of algebraic structure, the properties of operations, and reasoning to solve quadratic equations in one variable presented in a wide variety of forms (standard form, completing the square).
- Determine the conditions under which a quadratic equation has no solution, one real solution, or two real solutions.
- Use structure and reasoning to solve simple rational, radical, quadratic, exponential, polynomial, and absolute value equations in one variable.
- Solve systems of nonlinear equations in two variables.
- Fluently solve quadratic equations in one variable, either written as a quadratic expression in standard form equal to zero or completing the square.
- Use function notation to represent and interpret input/output pairs in terms of a context and points on the graph.
- Interpret the meaning of an input/output pair, constant, variable, factor, or term based on the context, particularly, seeing structure to solve the problem in its context.
- Identify, construct, and use quadratic or exponential functions to model a relationship between quantities and solve problems.

- Make connections between tabular, algebraic, and graphical representations of the function by (i) when given one representation, select another representation; (ii) identifying features of one representation given another representation, including maximum and minimum values of the function; (iii) determining how a graph is affected by a change to its equation, including a vertical shift or scaling of the graph.
- Factor a polynomial or simple rational function, represent and interpret input/output pairs in terms of a context and points on the graph; the solutions, intercepts, and key features in terms of a context;
- Identify the graph given an algebraic representation of the function or an algebraic representation given the graph (with or without a context).
- Interpret the effect of outliers on measures of center and spread.
- Estimate and interpret the slope and intercepts of the line of best fit for a given scatterplot in context.
- Represent data on two quantitative variables in graphs, tables, and scatterplots and describe how the variables are related.
- Fit linear, quadratic, and exponential functions to sets of data and use them to solve problems in the context of the data.
- Construct and interpret one- and two-way tables, tree diagrams, area models, and other representations to find relative frequency, probabilities, and conditional probabilities.
- Find probability and conditional probability in simple contexts.
- Use trigonometric ratios and the Pythagorean Theorem to solve multistep right triangle problems. Use concepts, theorems and proofs of congruence and similarity of triangles to solve problems.
- Know and apply relevant theorems such as (i) the vertical angle theorem; (ii) triangle similarity and congruence criteria; (iii) triangle angle sum theorem; and (iv) the relationship of angles formed when a transversal cuts parallel lines.
- Determine which statements may be required to prove certain relationships or to satisfy a given Theorem.

Students at Level 4, demonstrate a strong command of reasoning and the previous math relationship and skills in the first three levels. They can fluently apply linear and non-linear concepts (quadratic and exponential) to model and solve problems and make estimates that do not involve an exact number for either growth or decay. They can create, solve, and interpret non-linear functions and demonstrate how changes in parameters can affect their models, algebraically and graphically. They can apply right triangle and trigonometric concepts to solve real-world problems in a variety of contexts. Students can fluently solve multistep contextualized problems that integrate more than one standard. Overall, students effectively use a range of strategies and reasoning to solve a variety of problem types. They can solve unfamiliar problems by insightful, creative use of models. They can identify the logical assumptions in models, they can analyze, apply, interpret, and justify models with accurate, careful reasoning, and they can compare multiple plausible approaches to modeling in a setting and thoughtfully choose the most appropriate model.

In addition to the mathematical content and cognitive demand in Level 3, students will

- Solve problems in a variety of contexts using the Pythagorean Theorem; right triangle trigonometry; trigonometric ratios, and properties of special right triangles.
- Apply knowledge and understanding of the complex number system to add, subtract, multiply, and divide with complex numbers and solve problems
- Use definitions, properties, and theorems relating to circles and parts of circles, such as radii, diameters, tangents, angles, arcs, arc lengths, and sector areas, to solve problems.
- Create an equation of the form $(x - h)^2 + (y - k)^2 = r^2$ to represent a circle in the xy -plane.
- Describe how the change to the equation representing a circle in the xy -plane affects the graph of the circle or vice versa.
- Fluently, complete the square in an equation representing a circle to determine properties of the circle when it is graphed in the xy -plane, and use the distance formula in problems related to circles.

Multistate System of Student Assessment

Introduction to the Multistate Achievement Level Descriptors for Evidence-Based Reading and Writing

The Evidence-Based Reading and Writing (ERW) section of the redesigned SAT measures how well students demonstrate college and career readiness and how well students are performing relative to state-adopted college and career readiness ELA/literacy content standards. The ERW section of the SAT is designed to measure the attainment of essential ELA/literacy outcomes closely associated with college and career readiness and success as established by the best available evidence regarding necessary prerequisites for effective postsecondary entry without remediation. The ERW section is composed of two tests: (1) Reading and (2) Writing and Language.

The SAT **Reading Test** focuses on the skills and knowledge needed to demonstrate reading comprehension across a range of content areas, including US and world literature, history/social studies, and science passages; in passages of varying text types, including arguments, informative/explanatory texts, and narratives; in single and paired passages; in passages accompanied by one or more informational graphics, such as tables, graphs, and charts; and in passages across a defined range of text complexity from early high school to postsecondary-entry levels. The SAT **Writing and Language Test** focuses on a range of revision and editing skills performed in the context of multiparagraph passages in a variety of content areas, including history/social studies, the humanities, and science, as well as passages on career-related topics; passages of varying text types, including arguments, informative/explanatory texts, and nonfiction narratives; passages accompanied by one or more informational graphics, such as tables, graphs, and charts; and passages across a defined range of text complexity from early high school to postsecondary-entry levels. The domain descriptions found in tables 3 (Reading) and 4 (Writing and Language) offer more detail on the specific skills and knowledge consistently assessed on each of the two tests.

In the course of demonstrating their reading, writing, language, and reasoning achievement, students demonstrate achievement of skills and knowledge associated with four **subscores** and two **cross-test scores**.

Subscores

- *Command of Evidence*. The Command of Evidence (CoE) subscore is derived from selected test questions on both the Reading and the Writing and Language Tests. Students must demonstrate a range of related skills and knowledge, including the developed abilities to cite textual evidence (e.g., determining the best textual evidence for the answer to another question), analyze the use of evidence in arguments, and interpret data expressed quantitatively (Reading Test) and to improve the topic development of text by revising main ideas, improving support, sharpening focus, and using data from informational graphics (Writing and Language Test).

- *Words in Context.* The Words in Context (WiC) subscore is derived from selected test questions on both the Reading and the Writing and Language Tests. Students must demonstrate the developed abilities to determine word/phrase meanings in context and analyze word choice rhetorically (Reading Test) and to use language precisely and concisely, maintain or enhance style and tone, and improve syntax (Writing and Language Test).
- *Expression of Ideas.* The Expression of Ideas (EoI) subscore is derived from selected test questions on the Writing and Language Test. Students must demonstrate the developed abilities to improve the topic development, organization, and rhetorical language use in text.
- *Standard English Conventions.* The Standard English Conventions (SEC) subscore is derived from selected test questions on the Writing and Language Test. Students must demonstrate the developed ability to observe the conventions of Standard Written English sentence structure, usage, and punctuation.

Cross-Test Scores

- *Analysis in History/Social Studies.* The Analysis in History/Social Studies cross-test score is derived from selected test questions on the Reading, Writing and Language, and Math Tests. Students must demonstrate the developed abilities to analyze texts and make or enhance meaning in the history/social studies content area (Reading Test; Writing and Language Test) and to perform tasks grounded in history/social studies contexts (Math Test).
- *Analysis in Science.* The Analysis in Science cross-test score is derived from selected test questions on the Reading, Writing and Language, and Math Tests. Students must demonstrate the developed abilities to analyze texts and make or enhance meaning in the science content area (Reading Test; Writing and Language Test) and to perform tasks grounded in science contexts (Math Test).

The Evidence-Based Reading and Writing Achievement Level Descriptors (ALDs) are designed to describe college and career readiness on the SAT, in accordance with **18 DE Reg. 556¹**. The achievement level descriptors are text descriptions of the fundamental skills and knowledge demonstrated by students in each category of achievement.

¹ The highlighted text will serve as a placeholder. States will insert their own regulatory language.

Achievement Level Descriptors

Grade 11: Evidence-Based Reading and Writing

Achievement Level 4	Achievement Level 3	Achievement Level 2	Achievement Level 1
<p>The student has exceeded the achievement level and demonstrates a thorough understanding of the knowledge and skills needed for college and career readiness and achievement relative to the Common Core ELA/Literacy Content Standards.</p>	<p>The student has met the achievement level and demonstrates adequate understanding of the knowledge and skills needed for college and career readiness and achievement relative to the Common Core ELA/Literacy Content Standards.</p>	<p>The student partially meets the achievement level and demonstrates an incomplete understanding of the knowledge and skills needed for college and career readiness and achievement relative to the Common Core ELA/Literacy Content Standards.</p>	<p>The student has not met the achievement level and demonstrates a minimal understanding of the knowledge and skills needed for college and career readiness and achievement relative to the Common Core ELA/Literacy Content Standards.</p>

READING TEST

ACHIEVEMENT LEVEL 1

Students at this level demonstrate a minimal understanding by using passages and/or pairs of passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Table 1 on page 12). Students at this level will:

- Rarely identify information and ideas explicitly stated in a passage
- Rarely draw reasonable inferences from a passage
- Rarely extrapolate in a reasonable way from the information and ideas in a passage or apply information and ideas in a passage to a new, analogous situation
- Inconsistently cite textual evidence that best supports a given claim or point
- Identify an explicitly stated central idea or theme in a passage or determine an implicit central idea or theme from a passage with a single clear purpose
- Rarely identify a reasonable summary of a passage or of key information and ideas in a passage
- Determine a simple relationship between information, ideas, or people depicted in a passage (e.g., recognizing a basic comparison, contrast, or sequence)
- Determine the meaning of a relatively common word or phrase using clear context clues
- Rarely determine how the selection of specific words and phrases or the use of patterns of words and phrases shapes meaning and tone in a passage

- Identify basic text structures with a limited understanding of their impact
- Rarely identify the relationship between a particular part of a passage (e.g., a sentence) and the whole passage
- Inconsistently determine the point of view or perspective from which a passage is related or the influence this point of view or perspective has on content and style in a low to moderately complex passage
- Determine the main purpose of a low complexity passage
- Rarely identify a claim or counterclaim explicitly stated in an argument or determine an implicit claim or counterclaim from an argument
- Rarely assess an author’s reasoning for soundness
- Rarely assess how an author uses or fails to use evidence to support a claim or counterclaim
- Recognize a straightforward similarity or difference in a pair of moderately complex passages (e.g., in reading passages on the same topic, recognizing basic similarities and differences in how an event is depicted)
- Locate simple data or make a simple accurate interpretation of data in an informational graphic, such as a table, graph, or chart (e.g., comparing the size of two clearly labeled bars representing easy-to-interpret values on a bar graph)

ACHIEVEMENT LEVEL 2

Students at this level demonstrate an incomplete understanding by using passages and/or pairs of passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Table 1 on page 12). Students at this level will:

- Inconsistently identify information and ideas explicitly stated in a passage
- Inconsistently draw reasonable inferences from a passage
- Inconsistently extrapolate in a reasonable way from the information and ideas in a passage or apply information and ideas in a passage to a new, analogous situation
- Determine the best textual evidence for a conclusion when the evidence requires some interpretation or analysis
- Determine the central idea or theme of a complex passage
- Identify a simple summary of a passage or of key information and ideas in a passage
- Determine a basic relationship between information, ideas, or people depicted in a passage (e.g., establishing a cause-effect, comparison-contrast, or sequential relationship)
- Determine the meaning of a relatively common high-utility academic word or phrase in context; determine the literal meaning of a straightforward figurative expression in context
- Determine the main purpose or effect of an author's word choice in a complex passage or in a simpler passage when the purpose or effect is somewhat subtle (e.g., an author using words to convey a particular emotion)
- Describe basic text structures with a limited understanding of their impact
- Identify the relationship between a particular part of a passage (e.g., a sentence) and the whole passage

- Draw a straightforward reasonable inference about point of view or perspective in a complex passage (e.g., identifying a technique the author uses to shape point of view in a literary passage; distinguishing among the multiple perspectives in an informational passage)
- Determine the main purpose of a moderately complex passage
- Inconsistently identify a claim and counterclaim explicitly stated in an argument or determine an implicit claim or counterclaim from an argument
- Inconsistently assess an author’s reasoning for soundness
- Inconsistently assess how an author uses or fails to use evidence to support a claim or counterclaim
- Identify a similarity or difference in a pair of moderately complex passages (e.g., recognizing that a particular detail appears in one passage but not the other)
- Locate data or make an accurate interpretation of data in an informational graphic, such as a table, graph, or chart (e.g., drawing a valid conclusion based on an understanding of a bar graph's overall purpose; summarizing a clear trend from several data points)

ACHIEVEMENT LEVEL 3

Students at this level demonstrate an adequate understanding by using passages and/or pairs of passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Table 1 on page 12). Students at this level will:

- Identify information and ideas explicitly stated in or draw reasonable inferences from a moderately complex passage
- Extrapolate in a reasonable way from the information and ideas in a moderately complex passage or apply information and ideas in such a passage to a new, analogous situation
- Determine the best textual evidence for a conclusion when the evidence requires some interpretation or analysis and the conclusion may require making an inference
- Determine the central idea or theme of a complex passage that features several important ideas (e.g., making the most defensible interpretation of a literary passage that is subject to multiple interpretations; distinguishing the author's or narrator's main point or perspective from other points or perspectives represented in the passage)
- Identify an accurate summary of a passage or of key information and ideas in a passage
- Determine a relationship between information, ideas, or people in a moderately complex passage (e.g., establishing a cause-effect, comparison-contrast, or sequential relationship)
- Determine the meaning of a relatively uncommon high-utility academic word or phrase in context; determine the literal meaning of a moderately challenging figurative expression in context
- Determine the main purpose or effect of an author's word choice in a complex passage or in a simpler passage when the purpose or effect is fairly subtle or complex (e.g., an author using word play or parody)
- Determine the overall structure of a passage with an understanding of its impact
- Determine the main purpose of a particular part of a passage (e.g., a detail or a metaphor) in relation to the passage as a whole
- Draw a reasonable inference about point of view or perspective in a complex passage (e.g., identifying where point of view switches in a literary passage; distinguishing among conflicting perspectives in an informational passage)

- Determine the main purpose of a complex passage or that of one of its paragraphs
- Determine a claim or counterclaim in a complex argument
- Assess an author’s reasoning for soundness in a moderately complex argument; analyze an argumentative technique or flaw (e.g., an author using weak reasoning in support of a claim)
- Assess how an author uses or fails to use evidence to support a claim or counterclaim in a moderately complex argument
- Synthesize information and ideas from paired, moderately complex passages (e.g., authors’ positions)
- Make an accurate, somewhat subtle or complex interpretations of data in an informational graphic, such as a table, graph, or chart (e.g., comparing results in terms of two variables; recognizing an implication of the values represented on a table); draw a supportable connection between a graphic and its accompanying passage (e.g., characterizing a broad trend exhibited in a graph using the concepts and language of the passage)

ACHIEVEMENT LEVEL 4

Students at this level demonstrate a thorough understanding by using passages and/or pairs of passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Table 1 on page 12). Students at this level will:

- Identify information and ideas explicitly stated in or draw reasonable inferences from a complex to highly complex passage
- Extrapolate in a reasonable way from the information and ideas in a complex to highly complex passage or apply information and ideas in such a passage to a new, analogous situation
- Determine the best textual evidence for a conclusion when the evidence is subtle, abstract, or figurative and the conclusion requires making one or more inferences
- Determine the central idea or theme of a highly complex passage
- Identify an accurate summary of a highly complex passage or of key information and ideas in such a passage
- Determine a subtle or complex relationship between information, ideas, or people in a highly complex passage
- Determine the meaning of an uncommon high-utility academic word or phrase in context, including an archaic usage found in a passage from an earlier time period; determine the literal meaning of subtle or complex figurative language in context
- Determine the main purpose or effect of an author's word choice in a highly complex passage or in a simpler passage when the purpose or effect is subtle or highly complex (e.g., the author establishing meaning chiefly through tone via understatement, exaggeration, or sarcasm)
- Analyze the overall structure of a passage with a clear understanding of its impact
- Determine the main purpose of a particular part of a passage in relation to the passage as a whole when the purpose is subtle or complex (e.g., the author using rhetorical questions to indicate self-evident truths)

- Draw a nuanced inference about point of view or perspective in a complex to highly complex passage (e.g., tracing a subtle shift in point of view in a literary passage; associating particular opinions with the individuals who hold them in an informational passage)
- Determine the main purpose of a highly complex passage or of a particular section of such a passage (e.g., two or more paragraphs)
- Identify a claim or counterclaim explicitly stated in a highly complex argument or determine an implicit claim or counterclaim from such an argument
- Assess an author’s reasoning for soundness in a complex to highly complex argument; analyze a subtle argumentative technique or flaw (e.g., an author using biased characterizations in support of a claim)
- Assess how an author uses or fails to use evidence to support a claim or counterclaim in a complex to highly complex argument
- Synthesize information and ideas from paired, complex to highly complex passages (e.g., subtle or complex comparisons)
- Make an accurate subtle or complex interpretation of data in an informational graphic, such as a table, graph, or chart (e.g., comparing results in terms of three or more variables; determining which bars on a bar graph can reasonably be considered part of an overarching category); draw a subtle or complex supportable connection between a graphic and its accompanying passage (e.g., summarizing the results displayed in a table using the concepts and language of the passage)

WRITING AND LANGUAGE TEST

ACHIEVEMENT LEVEL 1

Students at this level demonstrate a minimal understanding by using passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Tables 1 and 2 on pages 12–13).

Students at this level will:

- Use supporting information to achieve a simple purpose (e.g., providing a short list of examples)
- Delete information that is obviously irrelevant to the main focus of a paragraph or passage (e.g., eliminating a detail that has no clear relationship to a passage's topic)
- Order the sentences in a paragraph to achieve a simple purpose (e.g., grouping related information together; establishing a basic chronology)
- Introduce a paragraph that has a clear, well-defined focus
- Use a transitional word or phrase to establish a simple logical relationship between sentences (e.g., indicating sharp contrast)
- Make an effective word or phrase choice in a straightforward situation (e.g., using a common but still appropriate expression instead of an awkward or meaningless one)
- Eliminate obvious wordiness or redundancy (e.g., removing repetition within a short phrase)
- Combine sentences in a relatively simple way (e.g., making a second sentence into a relative clause of the first) or to achieve a relatively simple purpose (e.g., eliminating obvious awkwardness or repetition)

- Form conventional, complete sentences, recognizing and correcting a clear and substantial disruption in structure (e.g., eliminating an obvious comma splice; correcting a lack of parallelism in a simple series of words; replacing a nonstandard relative pronoun with a standard one)
- Recognize and correct an obviously inappropriate shift in verb tense (e.g., the use of present tense when the context clearly calls for past tense)
- Maintain subject-verb or pronoun-antecedent agreement in a straightforward situation (e.g., ensuring agreement between subject and verb when the number of the subject is clear and the subject and verb appear close together in the sentence)
- Use conventional expression in a straightforward situation (e.g., recognizing and correcting a nonsensical expression; choosing a preposition that establishes a logical relationship [e.g., *with*, *for*]; selecting appropriately between common words that are frequently confused, such as *to* and *too*)
- Distinguish between singular and plural possessive nouns and between plural and possessive nouns
- Use commas to set off a simple nonrestrictive element (e.g., a phrase describing the person just named)
- Eliminate obviously unnecessary and disruptive punctuation (e.g., between verb and direct object)

ACHIEVEMENT LEVEL 2

Students at this level demonstrate an incomplete understanding by using passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Tables 1 and 2 on pages 12–13). Students at this level will:

- Clarify an aspect of the structure of a paragraph or passage (e.g., using a phrase to set up examples that follow in subsequent sentences)
- Use supporting information to achieve a straightforward purpose (e.g., providing a cause for an effect; offering direct support for a claim)
- Delete information that is clearly irrelevant to a paragraph or passage (e.g., eliminating a detail that interrupts an explanation or sequence or that obviously digresses from the main topic)
- Use a general understanding of an informational graphic, such as a table, graph, or chart, to revise a passage (e.g., drawing on knowledge of what a graph's bars represent to improve the accuracy of a passage's description of the graph)
- Order the sentences in a paragraph to achieve a straightforward purpose (e.g., placing a supporting detail immediately after a sentence that makes a claim)
- Introduce or conclude a passage based on a general understanding of the passage's content and purpose (e.g., adding a conclusion that restates the passage's main claim)
- Use a transitional word or phrase to establish a straightforward logical relationship between sentences (e.g., indicating sequence or contrast; introducing a definition)
- Make an effective word or phrase choice based on vocabulary knowledge and an understanding of the context (e.g., recognizing when a particular word is or is not commonly used to describe a person or object)
- Eliminate wordiness or redundancy within a sentence (e.g., recognizing when adjectives with the same meaning or very similar meanings, such as *fast* and *rapid*, are used to describe the same thing)

- Maintain a basic consistency in style and tone within a passage (e.g., revising language that is clearly too colloquial or formal for the context)
- Combine sentences in a straightforward way (e.g., making a second sentence into a prepositional phrase of the first) or to achieve a straightforward purpose (e.g., establishing a logical arrangement of sentence elements)
- Form conventional, complete sentences, recognizing and correcting a disruption in structure (e.g., eliminating an obvious, rhetorically inappropriate fragment; maintaining parallelism in a simple series of phrases; establishing a clear relationship between an introductory and main clause)
- Determine appropriate verb tense or pronoun person and number on the basis of an understanding of the context (e.g., correcting an inappropriate shift from past tense to present or past perfect tense; making a needed shift from past to present tense to signal a change in time frame; correcting an inappropriate shift from third person *they* to second person *you*)
- Recognize and correct an obviously vague or ambiguous pronoun (e.g., replacing a pronoun without a clear antecedent with the appropriate noun)
- Maintain subject-verb or pronoun-antecedent agreement in a somewhat challenging situation (e.g., ensuring agreement between subject and verb when a short phrase intervenes)
- Use conventional expression in a somewhat challenging situation (e.g., choosing the preposition that appropriately completes a phrasal verb; selecting appropriately between less-common words that are frequently confused, such as *effect* and *affect*)
- Distinguish among singular, singular possessive, plural, and plural possessive nouns
- Appropriately punctuate items in a series (e.g., a three-item series of nouns with accompanying adjectives)
- Use punctuation to set off a nonrestrictive element (e.g., an interrupting phrase); eliminate punctuation inappropriately setting off a simple restrictive element (e.g., a job title that precedes a person's name)
- Eliminate unnecessary punctuation in a somewhat challenging situation (e.g., between noun and preposition; between verb and a clause serving as its object)

ACHIEVEMENT LEVEL 3

Students at this level demonstrate an adequate understanding by using passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Tables 1 and 2 on pages 12–13).

Students at this level will:

- Establish and clarify the structure of a paragraph or passage (e.g., adding a sentence to frame a paragraph's discussion or to present a claim that the paragraph subsequently supports)
- Use supporting information to develop a point or claim logically (e.g., offering a specific, relevant example; using a quotation that clarifies a concept or observation)
- Sharpen the focus of a paragraph or passage by making a careful decision about adding, revising, or deleting information (e.g., eliminating material that is broadly relevant to a topic but that is poorly placed or integrated)
- Locate or accurately interpret data in an informational graphic, such as a table, graph, or chart, to revise a passage (e.g., identifying the value in a table that is associated with a particular condition;

distinguishing between accurate and inaccurate interpretations and between relevant and irrelevant information)

- Order the sentences in a paragraph to address a critical issue of logic or cohesion (e.g., adding a sentence to fill a discernible gap in a chronological sequence; repositioning a sentence to provide a needed transition between ideas)
- Introduce or conclude a passage based on an understanding of the passage's content and purpose (e.g., ensuring that the conclusion offers an adequate sense of closure; achieving a particular rhetorical aim, such as suggesting implications of the findings discussed in the passage)
- Use a transitional word, phrase, clause, or sentence to establish a logical relationship between sentences or paragraphs (e.g., signaling a shift in emphasis or focus)
- Make a nuanced word or phrase choice based on well-developed vocabulary knowledge and a well-developed understanding of the context (e.g., distinguishing among relatively uncommon words that have similar denotations but differing connotations or uses)
- Eliminate relatively subtle wordiness or redundancy within a sentence or between sentences (e.g., recognizing when information overexplains a concept already made clear and correcting accordingly; deleting repetition involving fairly sophisticated language)
- Make a careful decision about style and tone in a passage based on an understanding of the context (e.g., revising language that is too colloquial or formal in a fairly challenging context; achieving a particular rhetorical aim, such as establishing a particular sentence pattern or choosing language that sets a contextually appropriate mood)
- Combine sentences to accomplish a relatively subtle purpose (e.g., inserting a conjunction to establish a logical relationship; blending elements of two sentences to improve the logic and flow of ideas)
- Form conventional, complete sentences, recognizing and correcting a relatively subtle disruption in structure (e.g., eliminating a rhetorically inappropriate fragment created by the use of a semicolon; maintaining parallelism in a series of phrases; choosing or eliminating a conjunction based on an understanding of the syntax of a relatively sophisticated or long sentence; correcting an obvious dangling modifier)
- Determine appropriate verb tense and mood or pronoun person and number on the basis of a well-developed understanding of the context (e.g., making a needed shift into conditional mood to suggest a possible but not certain outcome)
- Recognize and correct a vague or ambiguous pronoun based on an understanding of the context (e.g., replacing an ambiguous pronoun with a noun after a close reading to determine what the appropriate noun should be)
- Make careful distinctions among the possessive determiners *its* and *their*, the contractions *it's* and *they're*, and the adverb *there*
- Maintain subject-verb or pronoun-antecedent agreement in a challenging situation (e.g., ensuring agreement between subject and verb when a clause or multiple short phrases intervene and possibly suggest a different number for the verb than the subject warrants)
- Use conventional expression in a challenging situation (e.g., selecting appropriately between relatively uncommon words that are frequently confused, such as *discrete* and *discreet*)
- Make careful distinctions among singular, singular possessive, plural, and plural possessive nouns based on an understanding of the context (e.g., noting that the article *the* establishes that the noun it precedes is singular or singular possessive)
- Make a careful decision about how or whether to use punctuation to set off one or more sentence elements based on an understanding of the context (e.g., determining whether an element is restrictive or nonrestrictive through a close reading of the context and then punctuating or not

punctuating accordingly; using matching punctuation, such as two commas rather than a comma and a dash, to set off a nonrestrictive element)

- Eliminate unnecessary punctuation in a challenging situation (e.g., between a long subject and the predicate; after a word or phrase, such as *including*, that sets up a list of examples)

ACHIEVEMENT LEVEL 4

Students at this level demonstrate a thorough understanding by using passages in a specified range of text complexities from grades 9–10 to postsecondary entry (see Tables 1 and 2 on pages 12–13).

Students at this level will:

- Make a sophisticated decision relating to the structure of a paragraph or passage (e.g., using a clause to set up information when the content and language are complex and the linkage is subtle)
- Use supporting information to develop a point or claim logically on the basis of a thorough understanding of a challenging context (e.g., drawing on logic and an understanding of the context to indicate the last step in a complex sequence; including an example that is similar in content to one or more other examples in a paragraph)
- Sharpen the focus of a paragraph or passage by making a sophisticated decision about adding, revising, or deleting information (e.g., adding or retaining nonessential but relevant material because it enhances meaning and clarity)
- Accurately interpret, paraphrase, or summarize data in an informational graphic, such as a table, graph, or chart, to revise a passage (e.g., encompassing multiple data points in a single relevant general statement)
- Order the sentences in a paragraph to address a complex or subtle issue of logic or cohesion (e.g., deciding to reposition rather than delete a sentence that, when properly placed, improves the flow of ideas in a paragraph)
- Use a transitional word, phrase, clause, or sentence to establish a complex or subtle logical relationship between sentences or paragraphs; recognize when such a device is not needed or problematic (e.g., drawing on an understanding of the context to eliminate a word or phrase, such as *therefore*, that wrongly suggests a cause-effect relationship)
- Make a sophisticated word or phrase choice based on highly developed vocabulary knowledge and a thorough understanding of a challenging context (e.g., distinguishing among uncommon words that have similar denotations but differing connotations or uses when the distinctions are subtle)
- Eliminate subtle wordiness or redundancy within a sentence or between sentences and paragraphs (e.g., recognizing that a sophisticated-sounding but wordy expression is less effective than a simpler and more economical one; eliminating the second appearance of the same detail in successive paragraphs)
- Make a sophisticated decision about style and tone in a passage based on a thorough understanding of the context (e.g., achieving a subtle rhetorical aim, such as closely matching a sentence pattern already established in a passage)
- Combine sentences to accomplish a complex or subtle purpose (e.g., drawing on an understanding of the context to place a blended sentence's emphasis on its most important idea)
- Form conventional, complete sentences, recognizing and correcting a complex or subtle disruption in structure (e.g., ensuring the completeness of a sentence with an uncommon structure, such as a sentence containing a subject clause beginning with *that*; correcting minor and easily overlooked violations of parallelism, such as the omission of a preposition, in a series of phrases)
- Maintain subject-verb and pronoun-antecedent agreement in a complex situation (e.g., between subject and verb when the two are widely separated and when intervening text suggests a different number for the verb than the subject warrants)

- Use conventional expression in a complex situation (e.g., selecting appropriately between uncommon words that are frequently confused, such as *defuse* and *diffuse*)
- Use a semicolon to join two closely related independent clauses
- Use a colon to introduce a list or an elaboration (e.g., a noun phrase renaming a previously mentioned concept; an independent clause explaining a point introduced earlier in a sentence)

Table 1a: Text Complexity Definition

The redesigned SAT’s passages/passage pair represent a specified range of text complexity from grades 9–10 to postsecondary entry. Text complexity is defined as:

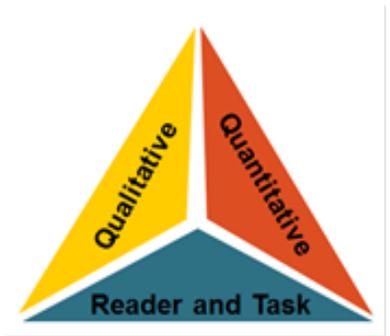
1. Quantitative measures – readability and other scores of text difficulty; often best measured by computer software.	<p style="text-align: center;">CCSS TEXT COMPLEXITY MODEL</p> 
2. Qualitative measures – levels of meaning, structure, and language conventionality and clarity, and knowledge demands; often best measured by an attentive human reader.	
3. Reader and task considerations – background knowledge of reader, motivation, interests, and complexity generated by tasks assigned; often best assessed by educators employing their professional judgment.	

Table 1b: Text Complexity Levels Referenced in the Redesigned SAT ALDs

Level of Text	Description
Low Complexity Text	These are texts whose information and ideas as well as structure, purpose, and language are relatively simple and direct and require relatively little analysis.
Moderately Complex Text	These are texts whose information and ideas as well as structure, purpose, and language require some analysis.
Complex Text	These are texts that can be difficult to understand at first because the information, ideas, structure, purpose, and language may be difficult or unfamiliar.
Highly Complex Text	These are texts that can be challenging even for very skilled readers. Often, the information, ideas, structure, purpose, and language are very difficult or unfamiliar.

Table 2: CCSS Language Progressive Skills Standards

Students advancing through the grades are expected to meet each year’s grade-specific standards and retain or further develop skills and understandings mastered in preceding grades. The following skills, marked with an asterisk (*) in Language standards 1–3, are particularly likely to require continued attention in higher grades as they are applied to increasingly sophisticated writing and speaking (see p. 56 of the CCSS).

- L.3.1f.** Ensure subject-verb and pronoun-antecedent agreement.
- L.3.3a.** Choose words and phrases for effect.
- L.4.1f.** Produce complete sentences, recognizing and correcting inappropriate fragments and run-ons.
- L.4.1g.** Correctly use frequently confused words (e.g., *to/too/two*; *there/their*)
- L.4.3a.** Choose words and phrases to convey ideas precisely.*
- L.4.3b.** Choose punctuation for effect.
- L.5.1d.** Recognize and correct inappropriate shifts in verb tense.
- L.5.2a.** Use punctuation to separate items in a series.†
- L.6.1c.** Recognize and correct inappropriate shifts in pronoun number and person.
- L.6.1d.** Recognize and correct vague pronouns (i.e., ones with unclear or ambiguous antecedents).
- L.6.1e.** Recognize variations from standard English in their own and others’ writing and speaking, and identify and use strategies to improve expression in conventional language.
- L.6.2a.** Use punctuation (commas, parentheses, dashes) to set off nonrestrictive/parenthetical elements.
- L.6.3a.** Vary sentence patterns for meaning, reader/listener interest, and style.‡
- L.6.3b.** Maintain consistency in style and tone.
- L.7.1c.** Place phrases and clauses within a sentence, recognizing and correcting misplaced and dangling modifiers.
- L.7.3a.** Choose language that expresses ideas precisely and concisely, recognizing and eliminating wordiness and redundancy.
- L.8.1d.** Recognize and correct inappropriate shifts in verb voice and mood.
- L.9–10.1a.** Use parallel structure

* Subsumed by L.7.3a
† Subsumed by L.9–10.1a
‡ Subsumed by L.11–12.3a

Table 3: SAT Reading Test Domain²

Content Dimension	Description
Text Complexity	The passages/pair on the SAT Reading Test represent a specified range of text complexities from grades 9–10 to postsecondary entry.
Information and Ideas	These questions focus on the informational content of text.
Reading closely	These questions focus on the explicit and implicit meaning of text and on extrapolating beyond the information and ideas in a text.
<i>Determining explicit meanings</i>	The student will identify information and ideas explicitly stated in text.
<i>Determining implicit meanings</i>	The student will draw reasonable inferences and logical conclusions from text.
<i>Using analogical reasoning</i>	The student will extrapolate in a reasonable way from the information and ideas in a text or apply information and ideas in a text to a new, analogous situation.
Citing textual evidence	The student will cite the textual evidence that best supports a given claim or point.
Determining central ideas and themes	The student will identify explicitly stated central ideas or themes in text and determine implicit central ideas or themes from text.
Summarizing	The student will identify a reasonable summary of a text or of key information and ideas in text.
Understanding relationships	The student will identify explicitly stated relationships or determine implicit relationships between and among individuals, events, or ideas (e.g., cause-effect, comparison-contrast, sequence).
Interpreting words and phrases in context	The student will determine the meaning of words and phrases in context.
Rhetoric	These questions focus on the rhetorical analysis of text.
Analyzing word choice	The student will determine how the selection of specific words and phrases or the use of patterns of words and phrases shapes meaning and tone in text.
Analyzing text structure	These questions focus on the overall structure of a text and on the relationship between a particular part of a text and the whole text.
<i>Analyzing overall text structure</i>	The student will describe the overall structure of a text.
<i>Analyzing part-whole relationships</i>	The student will analyze the relationship between a particular part of a text (e.g., a sentence) and the whole text.
Analyzing point of view	The student will determine the point of view or perspective from which a text is related or the influence this point of view or perspective has on content and style.
Analyzing purpose	The student will determine the main or most likely purpose of a text or of a particular part of a text (typically, one or more

² NOTE: Table 3 will be removed from this document after the standards setting.

	paragraphs).
Analyzing arguments	These questions focus on analyzing arguments for their content and structure.
<i>Analyzing claims and counterclaims</i>	The student will identify claims and counterclaims explicitly stated in text or determine implicit claims and counterclaims from text.
<i>Assessing reasoning</i>	The student will assess an author’s reasoning for soundness.
<i>Analyzing evidence</i>	The student will assess how an author uses or fails to use evidence to support a claim or counterclaim.
Synthesis	These questions focus on synthesizing multiple sources of information.
Analyzing multiple texts	The student will synthesize information and ideas from paired texts. (Note: All of the skills listed above may be tested with either single or paired passages.)
Analyzing quantitative information	The student will analyze information presented quantitatively in such forms as graphs, tables, and charts and/or relate that information to information presented in text.

Table 4: SAT Writing and Language Test Domain³

Content Dimension	Description
Text Complexity	The passages on the SAT Writing and Language Test represent a specified range of text complexities from grades 9–10 to postsecondary entry.
Expression of Ideas	These questions focus on revision of text for topic development, accuracy (consistency between text and graphic[s]), logic, cohesion, and rhetorically effective use of language.
Development	These questions focus on revising text in relation to rhetorical purpose. (Prior knowledge of the topic is not assessed, though consistency of the material within a passage may be.)
<i>Proposition</i>	The student will add, revise, or retain central ideas, main claims, counterclaims, topic sentences, and the like to structure text and convey arguments, information, and ideas clearly and effectively.
<i>Support</i>	The student will add, revise, or retain information and ideas (e.g., details, facts, statistics) intended to support claims or points in text.
<i>Focus</i>	The student will add, revise, retain, or delete information and ideas in text for the sake of relevance to topic and purpose.
<i>Quantitative information</i>	The student will relate information presented quantitatively in such forms as graphs, charts, and tables to information presented in text.
Organization	These questions focus on revision of text to improve the logic

³ Table 4 will be removed from this document after the standards setting.

	and cohesion of text at the sentence, paragraph, and whole-text levels.
<i>Logical sequence</i>	The student will revise text as needed to ensure that information and ideas are presented in the most logical order.
<i>Introductions, conclusions, and transitions</i>	The student will revise text as needed to improve the beginning or ending of a text or paragraph and to ensure that transition words, phrases, or sentences are used effectively to connect information and ideas.
Effective language use	These questions focus on revision of text to improve the use of language to accomplish particular rhetorical purposes.
<i>Precision</i>	The student will revise text as needed to improve the exactness or content appropriateness of word choice.
<i>Concision</i>	The student will revise text as needed to improve the economy of word choice (i.e., to eliminate wordiness and redundancy).
<i>Style and tone</i>	The student will revise text as necessary to ensure consistency of style and tone within a text or to improve the match of style and tone to purpose.
<i>Syntax</i>	The student will use various sentence structures to accomplish needed rhetorical purposes.
Standard English Conventions	These questions focus on editing text to ensure conformity to the conventions of Standard Written English sentence structure, usage, and punctuation.
Sentence structure	These questions focus on editing text to correct problems in sentence formation and inappropriate shifts in construction within and between sentences.
<i>Sentence formation</i>	These questions focus on editing text to correct problems with forming grammatically complete and standard sentences.
Sentence boundaries	The student will recognize and correct grammatically incomplete sentences (e.g., rhetorically inappropriate fragments and run-ons).
Subordination and coordination	The student will recognize and correct problems in coordination and subordination in sentences.
Parallel structure	The student will recognize and correct problems in parallel structure in sentences.
Modifier placement	The student will recognize and correct problems in modifier placement (e.g., misplaced or dangling modifiers).
<i>Inappropriate shifts in construction</i>	These questions focus on editing text to correct inappropriate shifts in verb tense, voice, and mood and pronoun person and number.
Verb tense, mood, and voice	The student will recognize and correct inappropriate shifts in verb tense, voice, and mood within and between sentences.
Pronoun person and number	The student will recognize and correct inappropriate shifts in pronoun person and number within and between sentences.
Conventions of Usage	These questions focus on editing text to ensure conformity to the conventions of Standard Written English usage.
<i>Pronouns</i>	These questions focus on the proper use of pronouns.
Pronoun clarity	The student will recognize and correct pronouns with unclear or ambiguous antecedents.

<i>Possessive determiners</i>	The student will recognize and correct cases in which possessive determiners (<i>its, your, their</i>), contractions (<i>it's, you're, they're</i>), and adverbs (<i>there</i>) are confused with each other.
<i>Agreement</i>	These questions focus on ensuring grammatical agreement.
Pronoun-antecedent agreement	The student will recognize and correct lack of agreement between pronoun and antecedent.
Subject-verb agreement	The student will recognize and correct lack of agreement between subject and verb.
Noun agreement	The student will recognize and correct lack of agreement between nouns.
<i>Frequently confused words</i>	The student will recognize and correct instances in which a word or phrase is confused with another (e.g., <i>accept/except, allusion/illusion</i>).
<i>Logical comparison</i>	The student will recognize and correct cases in which unlike terms are compared.
<i>Conventional expression</i>	The student will recognize and correct cases in which a given expression is inconsistent with Standard Written English.
Conventions of Punctuation	These questions focus on editing text to ensure conformity to the conventions of Standard Written English punctuation.
<i>End-of-sentence punctuation</i>	The student will recognize and correct inappropriate uses of ending punctuation in cases in which the context makes the intent clear.
<i>Within-sentence punctuation</i>	The student will correctly use and recognize and correct inappropriate uses of colons, semicolons, and dashes to indicate sharp breaks in thought within sentences.
<i>Possessive nouns and pronouns</i>	The student will recognize and correct inappropriate uses of possessive nouns and pronouns as well as differentiate between possessive and plural forms.
<i>Items in a series</i>	The student will correctly use and recognize and correct inappropriate uses of punctuation (commas and sometimes semicolons) to separate items in a series.
<i>Nonrestrictive and parenthetical elements</i>	The student will correctly use punctuation (commas, parentheses, dashes) to set off nonrestrictive and parenthetical sentence elements as well as recognize and correct cases in which restrictive or essential sentence elements are inappropriately set off with punctuation.
<i>Unnecessary punctuation</i>	The student will recognize and correct cases in which unnecessary punctuation appears in a sentence.

Appendix D: Operational Rating Sheet (Example only, not actual sheet used)

	A	B	C	D	E	F	G	H	I	J
1	Use the pull down menu found by clicking into each cell to select the percentage of examinees from a sample of 100 borderline									
2	examinees at each cut score that would be expected to answer the item correctly. For example if you think 50 out of 100 borderline									
3	examinees at the 4/5 cut would answer the item 1 correctly then you would select 50% for the first cell in the purple column. Once all									
4	your ratings for an item have been entered, if any red cells remain please take another look to be sure all cells to the right have an equal or higher value than the cell immediately to the left of it. Values at or below chance for an individual item are highlighted in yellow and should be used sparingly.									
5						Average Rating				
6	Item	1/2 cut	2/3 cut	3/4 cut	4/5 cut	1/2 cut	2/3 cut	3/4 cut	4/5 cut	
7	1					0.00	0.00	0.00	0.00	
8	2									
9	3									
10	4									
11	5									
12	6									
13	7									
14	8									
15	In the cells below, enter the average score to one decimal place in increments									
16	of 0.5 on each free response question (FRQ) that you would expect to be									
17	achieved by 100 borderline students at that cut score.									
18	Item	1/2 cut	2/3 cut	3/4 cut	4/5 cut					
19	FRQ 1									
20	FRQ 2									
21										
22										
23	Multiple									
24	Choice Cut									
25		0.00	0.00	0.00	0.00					
26	Free Response									
27	Cut									
28		0.00	0.00	0.00	0.00					
29										

If any of the above values are highlighted in yellow it means the cut score you are recommending is currently below chance and you should take another look at the items you rated the lowest in that category to be sure the rating is appropriate.

When you have completed your ratings for this round there should be no red highlighting remaining. If any of the values above are highlighted in red it means that the overall cut score is at or below chance and must be raised so that it is above chance.

Appendix E: Training Evaluation Form

Panelist's ID#: _____

Training /Ready to Proceed Evaluation

SAT Evidence-Based Reading and Writing

The purpose of this form is to verify whether you understand the general purpose of the standard setting study and believe that you have received sufficient information and explanation to make your standard setting judgments.

	Yes	No
I understand the purpose of the standard setting study.		
I understand the steps I am to follow to make my standard setting judgments.		
I understand the concept of the borderline examinee.		
I am ready to complete my standard setting judgments.		

If you responded "No" to any of these statements, please indicate what additional information or explanations you need.

(Date)

(Signature)

(Print Name)

□

Appendix F: Round 1 Evaluation

Round 1 Evaluation – SAT Evidence-Based Reading and Writing Standard Setting

For questions 1-6, please indicate below the degree to which you agree with each of the following statements.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I understand the purpose of the study.				
2. The instructions and explanations provided by the facilitator were clear.				
3. The training in the standard setting method gave me the information I needed to complete my assignment.				
4. The ALDs that were developed prior to the meeting were accurate.				
5. I understand the concept of the borderline examinee.				
6. The ALDs helped me to determine how to rate each item.				

For questions 7-11, indicate how **influential** each of the following factors was in completing the item ratings.

	Not Influential	Influential	Very Influential
7. Completing the test before beginning the task.			
8. My perception of the difficulty of the items			
9. The ALDs			
10. The consequences of the test for students			
11. My experience with students in my classroom			

12. Please identify any additional factors you considered in making your Round 1 ratings that were not included above. _____

13. Do you have other comments that you would like to share at this time (use back of paper if needed)? _____

Appendix G: Final Evaluation Form

Final Evaluation Form for SAT Evidence-Based Reading and Writing

For questions 1-13, please indicate below the degree to which you agree with each of the following statements.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I understood the purpose of the study.				
2. The instructions and explanations provided by the facilitator were clear.				
3. The training on the standard setting method gave me the information I needed to complete my assignment.				
4. The ALDs that were developed prior to the meeting were accurate.				
5. I understood the concept of the borderline examinee.				
6. The ALDs helped me to determine how to rate each item.				
7. It was beneficial to have an opportunity for discussion and to review feedback between rounds.				
8. The impact data showing the percent of students expected to place into each category made a difference in how I rated the items in the next round.				
9. The opportunity to more than 1 round of ratings (i.e., round 2) helped me to feel more confident about my final ratings.				
10. I felt engaged in the process.				
11. I was comfortable sharing my ideas with the other panelists during the discussions.				
12. I am confident this standard setting process will produce fair cut scores.				
13. I would be comfortable defending this process to my peers.				

(OVER)

14. If you answered Disagree or Strongly Disagree to any of the above, please give us feedback on what could have been done differently to change this outcome. Please attach additional pages, if needed.

For questions 15-25 indicate how **influential** each of the following factors was in completing your ratings.

	Not Influential	Influential	Very Influential
15. Completing the test before beginning the task.			
16. My perception of the difficulty of the items			
17. The actual item difficulty provided for each item			
18. Distributions of students expected to earn each achievement level			
19. Small group discussion after Round 1			
20. Large group discussion after Round 1			
21. Discussion after Round 2			
22. The average ratings of other panelists			
23. The ALDs			
24. The consequences of the exam for students			
25. My experience with students in my classroom			

(Next Page)

26. Were there other factors that influenced your ratings in any of the three rounds that were not described above? If yes, please describe the other factors. Please attach additional pages, if needed.

For questions 27-35, first indicate how **useful** each of the following materials or procedures was in completing the exercises over the past couple of days.

	Not Useful	Useful	Very Useful
27. Taking the exam prior to beginning the task			
28. Practicing the procedure with real items prior to beginning the actual rating task			
29. Referencing the ALDs			
30. Small Group Discussion after round 1			
31. Large Group Discussion after round 1			
32. Test Overview			
33. Actual Item Difficulty Values			
34. Distribution of students earning each achievement level			
35. Large Group Discussion after round 2			

(OVER)

36. Can you think of other information that would have been useful to you during the standard setting process in helping you make your ratings? If yes, Please indicate what other information you would have found useful. Please attach additional pages, if needed.

For questions 37-44, please indicate how appropriate the amount of time was to complete the different components of the standard setting task.

	Too Little Time	About Right	Too Much Time
37. Taking the test			
38. Reviewing the ALDs			
39. Training on the rating task before Round 1			
40. Round 1 of the rating task			
41. Small Group Discussion after round 1			
42. Large Group Discussion after round 1			
43. Round 2 of the rating task			
44. Large Group Discussion after round 2			

45. Please share any comments that you may have on the items, standard setting process, meeting logistics, or other topic of interest. Attach additional pages, if needed.

Appendix H: Example Individual-Panelist Feedback Form

SAT ERW - Round 1 Results - Panelist 1

Multiple Choice Questions	Your Recommendation	Group Recommendation	Total Number of MC Items on Test
Number of MCQ items (out of 96) the student must get correct to be placed into Level 2:	49.85	44.125	96
Number of MCQ items (out of 96) the student must get correct to be placed into Level 3:	64.50	64.425	96
Number of MCQ items (out of 96) the student must get correct to be placed into Level 4:	79.00	82.425	96

Appendix I: Item-Level Feedback Sample

Multiple Choice Questions Feedback				
Item No.	Median Group Rating at Each Cut Score			Item Difficulty (Larger = Easier)
	1/2	2/3	3/4	
RD 1	40	65	90	0.71
RD 2	50	70	95	0.84
RD 3	32.5	52.5	82.5	0.73
RD 4	40	60	85	0.77
RD 5	30	60	85	0.56
RD 6	50	70	90	0.92
RD 7	35	65	85	0.66
RD 8	32.5	55	85	0.54
RD 9	60	80	95	0.56
RD 10	60	77.5	95	0.91
RD 11	40	62.5	85	0.61
RD 12	32.5	60	85	0.42
RD 13	30	60	80	0.68
RD 14	50	75	90	0.69
RD 15	37.5	60	82.5	0.55
RD 16	40	65	82.5	0.82
RD 17	40	60	80	0.84
RD 18	45	70	92.5	0.87
RD 19	32.5	65	87.5	0.87
RD 20	32.5	57.5	90	0.52
RD 21	30	50	77.5	0.59
RD 22	35	62.5	85	0.82
RD 23	55	77.5	92.5	0.90
RD 24	40	67.5	87.5	0.59
RD 25	40	65	85	0.61
RD 26	50	80	95	0.82
RD 27	37.5	60	85	0.61
RD 28	32.5	60	87.5	0.64
RD 29	50	70	90	0.89
RD 30	47.5	75	90	0.60
RD 31	37.5	60	80	0.30
RD 32	40	65	82.5	0.43
RD 33	35	62.5	85	0.76
RD 34	35	65	87.5	0.72
RD 35	50	75	92.5	0.69
RD 36	35	62.5	85	0.55
RD 37	52.5	80	95	0.61
RD 38	30	52.5	82.5	0.42
RD 39	30	52.5	80	0.52

Multiple Choice Questions Feedback				
Item No.	Median Group Rating at Each Cut Score			Item Difficulty (Larger = Easier)
	1/2	2/3	3/4	
RD 40	40	62.5	82.5	0.44
RD 41	40	62.5	87.5	0.50
RD 42	30	55	77.5	0.21
RD 43	45	70	90	0.51
RD 44	57.5	80	95	0.76
RD 45	42.5	67.5	85	0.73
RD 46	40	62.5	85	0.45
RD 47	52.5	72.5	90	0.79
RD 48	50	70	90	0.50
RD 49	50	77.5	90	0.67
RD 50	50	72.5	87.5	0.65
RD 51	37.5	70	90	0.68
RD 52	50	70	87.5	0.70
WL 1	50	70	95	0.68
WL 2	40	70	95	0.93
WL 3	47.5	70	90	0.79
WL 4	50	70	87.5	0.32
WL 5	37.5	70	85	0.64
WL 6	57.5	77.5	90	0.91
WL 7	60	80	95	0.70
WL 8	60	77.5	92.5	0.91
WL 9	50	72.5	90	0.72
WL 10	42.5	65	82.5	0.77
WL 11	30	57.5	82.5	0.32
WL 12	60	80	95	0.72
WL 13	50	70	90	0.10
WL 14	52.5	72.5	90	0.76
WL 15	45	70	90	0.49
WL 16	30	57.5	85	0.72
WL 17	52.5	72.5	90	0.89
WL 18	42.5	70	85	0.66
WL 19	50	72.5	90	0.91
WL 20	40	65	85	0.52
WL 21	57.5	75	92.5	0.42
WL 22	42.5	70	90	0.53
WL 23	50	70	85	0.75
WL 24	50	77.5	90	0.74
WL 25	50	72.5	92.5	0.40
WL 26	42.5	65	85	0.54

Appendix J: SAT Essay Standards Verification

In addition to cut scores on the SAT Math and Evidence-Based Reading and Writing Sections, Delaware and Maine desired a single cut score be established on the SAT essay to differentiate students with a minimal level of competency in productive writing from those without. This process began with an empirical analysis of student performance data on the SAT essay with those students scoring at or above the final Level 3 cut score of 480 on the SAT Evidence-Based Reading and Writing Section forming one group and those scoring below that point forming a second group. Logistic regression was performed to identify the essay total sum score (adding the three subscores together) consistent with a 0.50 probability of scoring 480 or higher. Combined data from Maine and Delaware was used in the analyses. Scores of zero were not included in the data set. Results were rounded up to the closest integer score and this value became the recommended cut score for the SAT essay. The following tables provide information on the results of the analyses conducted to produce the recommended cut score of a total sum score of 12 on the essay.

Table 1. Frequency Distribution of Essay Total Sum Score

ESSAY_TOTAL_SCORE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
6	1169	6.09	1169	6.09
7	887	4.62	2056	10.70
8	1112	5.79	3168	16.49
9	1256	6.54	4424	23.03
10	1752	9.12	6176	32.15
11	1762	9.17	7938	41.33
12	1846	9.61	9784	50.94
13	1505	7.84	11289	58.78
14	1549	8.06	12838	66.84
15	1443	7.51	14281	74.35
16	1364	7.10	15645	81.45
17	1319	6.87	16964	88.32
18	1082	5.63	18046	93.96
19	300	1.56	18346	95.52
20	324	1.69	18670	97.20
21	297	1.55	18967	98.75
22	85	0.44	19052	99.19
23	78	0.41	19130	99.60
24	77	0.40	19207	100.00

Table 2. Frequency Distribution of Assigned Scores on the Reading Subscore of the Essay

ESSAY_READING_SCORE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	1523	7.93	1523	7.93
3	2187	11.39	3710	19.32
4	5096	26.53	8806	45.85
5	4565	23.77	13371	69.62
6	4578	23.84	17949	93.45
7	1017	5.29	18966	98.75
8	241	1.25	19207	100.00

Table 3. Frequency Distribution of Assigned Scores on the Writing Subscore of the Essay

ESSAY_WRITING_SCORE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	2224	11.58	2224	11.58
3	2455	12.78	4679	24.36
4	4973	25.89	9652	50.25
5	3983	20.74	13635	70.99
6	4502	23.44	18137	94.43
7	868	4.52	19005	98.95
8	202	1.05	19207	100.00

Table 4. Frequency Distribution of Assigned Scores on the Analysis Subscore of the Essay

ESSAY_ANALYSIS_SCORE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	6950	36.18	6950	36.18
3	3415	17.78	10365	53.96
4	3969	20.66	14334	74.63
5	2626	13.67	16960	88.30
6	1636	8.52	18596	96.82
7	485	2.53	19081	99.34
8	126	0.66	19207	100.00

Table 5. Descriptive Statistics on Total Sum Score and Subscores for the Aggregated Data

Variable	N	Mean	Std Dev	Minimum	Maximum
ESSAY_TOTAL_SCORE	19207	12.65	3.93	6.00	24.00
ESSAY_READING_SCORE	19207	4.65	1.37	2.00	8.00
ESSAY_WRITING_SCORE	19207	4.49	1.44	2.00	8.00
ESSAY_ANALYSIS_SCORE	19207	3.51	1.48	2.00	8.00

Table 6. Descriptive Statistics on Total Sum Score and Subscores by State

STATE	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
DE	8078	ESSAY_READING_SCORE	8078	4.65	1.32	2.00	8.00
		ESSAY_WRITING_SCORE	8078	4.49	1.41	2.00	8.00
		ESSAY_ANALYSIS_SCORE	8078	3.54	1.46	2.00	8.00
		ESSAY_TOTAL_SCORE	8078	12.68	3.85	6.00	24.00
ME	11129	ESSAY_READING_SCORE	11129	4.65	1.40	2.00	8.00
		ESSAY_WRITING_SCORE	11129	4.50	1.46	2.00	8.00
		ESSAY_ANALYSIS_SCORE	11129	3.49	1.50	2.00	8.00
		ESSAY_TOTAL_SCORE	11129	12.64	3.99	6.00	24.00

Table 7. Logistic Regression Results for Varying Levels of Probability for the Aggregated Data and by State Using Total Sum Score and by Subscore.

STATE	ESSAY	Prob=0.30	Prob=0.40	Prob=0.50	Prob=0.55	Prob=0.60	Prob=0.65	Prob=0.70	Prob=0.75	Prob=0.80	Prob=0.85	Prob=0.90
OVERALL	TOTAL_SCORE	9.56	10.55	11.46	11.91	12.37	12.85	13.37	13.93	14.58	15.36	16.40
OVERALL	READING_SCORE	3.37	3.82	4.24	4.44	4.65	4.87	5.10	5.36	5.65	6.01	6.48
OVERALL	WRITING_SCORE	3.29	3.70	4.08	4.27	4.47	4.67	4.88	5.12	5.39	5.72	6.15
OVERALL	ANALYSIS_SCORE	2.16	2.58	2.96	3.15	3.35	3.55	3.77	4.01	4.28	4.61	5.05
DE	TOTAL_SCORE	10.16	11.17	12.11	12.57	13.04	13.53	14.05	14.63	15.29	16.09	17.16
DE	READING_SCORE	3.63	4.07	4.48	4.68	4.89	5.10	5.33	5.58	5.87	6.22	6.69
DE	WRITING_SCORE	3.50	3.93	4.32	4.51	4.71	4.91	5.13	5.37	5.65	5.98	6.42
DE	ANALYSIS_SCORE	2.40	2.85	3.25	3.45	3.66	3.87	4.10	4.35	4.64	4.99	5.46
ME	TOTAL_SCORE	9.15	10.11	10.99	11.42	11.87	12.33	12.82	13.37	13.99	14.75	15.75
ME	READING_SCORE	3.18	3.63	4.05	4.25	4.46	4.68	4.91	5.17	5.46	5.82	6.29
ME	WRITING_SCORE	3.14	3.54	3.91	4.10	4.28	4.48	4.69	4.92	5.18	5.50	5.93
ME	ANALYSIS_SCORE	2.02	2.41	2.76	2.94	3.12	3.30	3.50	3.72	3.98	4.28	4.69

Results in Tables 1-7 were shared with both Delaware and Maine, along with the College Board recommendation that a total sum score of 12 be the recommended cut score on the essay. After review, both states agreed to move forward with the recommendation of a total sum score of 12 on the essay. On August 5, 2016, a group of 6 subject matter experts (SMEs) met for a half day to review the recommended cut score, student work samples at each total sum score, and impact data. This group contained 4 SMEs from Delaware and 2 SMEs from Maine. Table 8 provides a summary of the demographics of this group of six.

Table 8. Demographics of Subject Matter Experts

Totals	Counts
Female	5
Male	1
American Indian or Alaska Native	1
Asian, Asian American, or Pacific Islander	0
Black or African American	2
Mexican or Mexican American	0
Puerto Rican	0
Other Hispanic, Latino, or Latin American	0
White	3
Other	0
1-3 years	0
4-6 years	2
7-12 years	0
12+ years	4
Teach at a High School	4
Teach at College	0
Teach both College and High School	2
Teach at Other	0
Undergraduate Degree	0
Master's Degree	3
Specialist's Degree	1
Doctoral Degree	1
Other Education Level	0

During the standards verification meeting, the SMEs were provided with an overview of the SAT essay, the rubrics, and the scoring process by members of the College Board Assessment Design and Development team. The recommended cut score on the essay of a Total Sum Score of 12 was introduced along with an explanation of how this score was identified. Additionally, the concept of the score being compensatory, meaning no minimum value is specified on any subscore and all combinations that sum to 12 or higher would be considered to have met the expectations, was explained and discussed.

The SMEs were provided with 30 student essay responses grouped into two packets, those that meet or exceed expectations and those that do not meet expectations. The essays represented scores from across the score range and were selected because they were frequently observed score patterns at that Total Sum Score point. Where possible, preference was given to patterns/papers where the two raters were in exact agreement. No scores were visible on the actual student responses. Panelists were asked to review each student response and indicate whether they agree the paper is correctly placed in the Meets or Exceeds Expectations packet or should be in the Does Not Meet Expectations packet (and vice versa). Within the two packets, responses were ordered from highest total sum score to lowest total sum score but this was not shared with the panelists until after their independent review when all Total Sum Scores and Subscores were revealed for each sample. Figures 1 and 2 provide examples of the materials shared with the panelists during and following the response sample review.

Discussion following the independent review of student response samples was very positive and panelists were in agreement that the majority of the essay response samples were in the correct packet. Samples 16 to 21 were identified as borderline and once panelists saw the scores for those samples they felt it made sense. Some discussion focused on the level of rigor and saying it was acceptable to have students receive a 2 on each of the three subscores and meet expectations. Panelists were a little uncomfortable with that, but felt it was reasonable for now given the different nature of the task required for the SAT essay compared to what students had been prepared for historically. Panelists expressed a desire to revisit the SAT essay standard in a few years when students have had more opportunity to become familiar with the task required.

As part of the discussion impact data was provided in the form of the Mean and Standard Deviation on the Total Sum Score for the essay for the aggregated data combining Delaware and Maine (Mean=12.65, SD = 3.93), and disaggregated by state (Delaware: Mean=12.68, SD=3.85; Maine: Mean=12.64, SD=3.99). The percentages of students from the aggregated data that would be expected to Meet or Exceed the Standard (58.67%) and those that would not meet the standard (41.33%) were also shared. Once the discussion concluded, panelists were asked to complete a recommendation form as shown in Figure 3. All panelists indicated on all 3 questions that the level of rigor is appropriate, the impact met their expectations, and they would make no change to the recommended cut score of 12. Comments provided on the recommendation forms are included below:

I would recommend that the cut score be revisited in a few years once CCSS has been better implemented to prepare the students.
After this morning's conversation, I believe 12 is a good starting point. However, given that a student could earn all partially meets and still "meet the standard" we should be working towards a higher cut score in which students achieve 3" at least at some level.
Perhaps revisit this cut score in 3 years after more schools have had a chance to undergo PBE curriculum work to shift more aggressively to standards based instruction.
It may be valuable to return to this cut score for review in 5 years. High stakes opportunities prior would be valuable.
The cut score seems appropriate given the sample size; however, I would caution that kids at the borderline for not meeting the standard (i.e., 10.5/11) may be erroneously categorized.

Following completion of the recommendation form, panelists completed a final evaluation form. The results of the final evaluation form are summarized in Table 9. After completion of the final evaluation form, panelists were thanked for their participation and dismissed.

Figure 1. Record Form for Response Sample Review

ID# _____

SAT Essay Standards Verification Record

Independently review each essay. If you agree the essay is in the correct group, place a check mark (✓) on the blank. If you do not agree the essay is in the correct group, place a dash (-) on the blank.

Meets or Exceeds Standard	Does Not Meet Standard
Essay 1. _____	Essay 20. _____
2. _____	21. _____
3. _____	22. _____
4. _____	23. _____
5. _____	24. _____
6. _____	25. _____
7. _____	26. _____
8. _____	27. _____
9. _____	28. _____
10. _____	29. _____
11. _____	30. _____
12. _____	
13. _____	
14. _____	
15. _____	
16. _____	
17. _____	
18. _____	
19. _____	

Figure 2. Scores for Response Samples in Each Packet

Alignment of Essay Number to Subscores and Sum Total				
Essay ID	Essay Reading	Essay Analysis	Essay Writing	Sum Total
1	8	8	8	24
2	8	7	8	23
3	8	7	7	22
4	7	7	7	21
5	7	6	7	20
6	7	6	6	19
7	6	6	6	18
8	6	5	6	17
9	4	6	6	16
10	6	4	6	16
11	5	5	5	15
12	6	4	6	16
13	6	2	6	14
14	5	4	5	14
15	5	4	4	13
16	5	3	5	13
17	4	2	6	12
18	6	2	4	12
19	4	4	4	12
20	6	2	3	11
21	4	4	3	11
22	4	2	5	11
23	4	3	4	11
24	2	4	4	10
25	3	4	3	10
26	4	2	4	10
27	5	2	2	9
28	4	2	3	9
29	4	2	2	8
30	2	2	2	6

Figure 3. Panelist Recommendation Form

Standards Verification Record Form – SAT Essay

The recommended cut score for the SAT essay is a sum total of 12. With this recommended cut score, 58.67% of students from Maine and Delaware are expected to meet or exceed the standard.

Please mark the boxes below that correspond to the statements that best characterize your opinions regarding this percentage and the recommended cut score.

1. Based on your review of student work samples, is the recommended level of rigor expected to Meet or Exceed the Standard on the SAT Essay appropriate?
 Yes, the level of rigor expected to earn the recommended cut score of 12 is appropriate.
 No, the expected level of rigor is too low.
 No, the expected level of rigor is too high.
2. Based on your experience with students and knowledge of the SAT essay rubric and scoring process, were you surprised by the percentage of students expected to Meet or Exceed the Standard on the SAT essay with the recommended cut score of 12?
 No, this matches my expectation
 Yes, I expected more students to Meet or Exceed the Standard.
 Yes, I expected fewer students to Meet or Exceed the Standard.
3. Please mark the box corresponding to the response that indicates *how you would change the final cut score*.
 Make no change. I am satisfied with the recommended cut score of 12.
 I would like the cut score to be raised to represent a more rigorous standard.
 I would like the cut score to be lowered, the current recommendation is too rigorous.

Please provide any comments you would like to be considered by the DOE and State Board regarding the rigor of the recommended cut score and the impact on students before the cut score for the SAT essay is finalized. Feel free to use the back of the page, if necessary.

Table 9. Final Evaluation Form Results

Strongly Agree(4) to Strongly Disagree(1)		
1	I understood the purpose of the study.	4.0
2	The facilitator explained things clearly.	4.0
3	I felt comfortable expressing my opinions.	4.0
4	The training on the essay, rubric, and scoring process provided the information I needed to make an informed decision.	3.8
5	I understood how the recommended cut score was identified.	3.8
6	Reviewing samples of student responses gave me confidence in the location of the recommended cut score.	3.8
7	The impact data was helpful to me in confirming whether the cut score is appropriate.	3.8
8	I am comfortable with the cut score recommended.	3.7
9	I would be comfortable defending this process to my peers.	3.6
Very Useful(3) to Not at all Useful (1)		
10	Reviewing the prompt, rubrics, and scoring process.	3.0
11	Reviewing student response samples.	3.0
12	Explanation of how the recommended cut score was identified.	2.8
13	Impact data.	2.8
14	Group discussions.	2.8
Very Influential (3) to Not Influential (1)		
15	My perception of the accessibility of the prompt to students.	3.0
16	The different patterns of scores a student could earn to reach the cut score.	2.7
17	My personal experiences with students.	2.8
18	Discussions with other panelists.	2.8
19	The impact data.	2.7
20	The importance of the test to the examinees.	2.0

Final Evaluation Form Comments:

I think it is important to make sure the audience understands the sample size & how reflective the samples are of their population of students. I am comfortable defending this process based on the type of statistical analyses used.
Given that a student could earn 2 in all categories and still meet the standard, this should be a starting point but we should work towards setting a higher cut score in the future.