

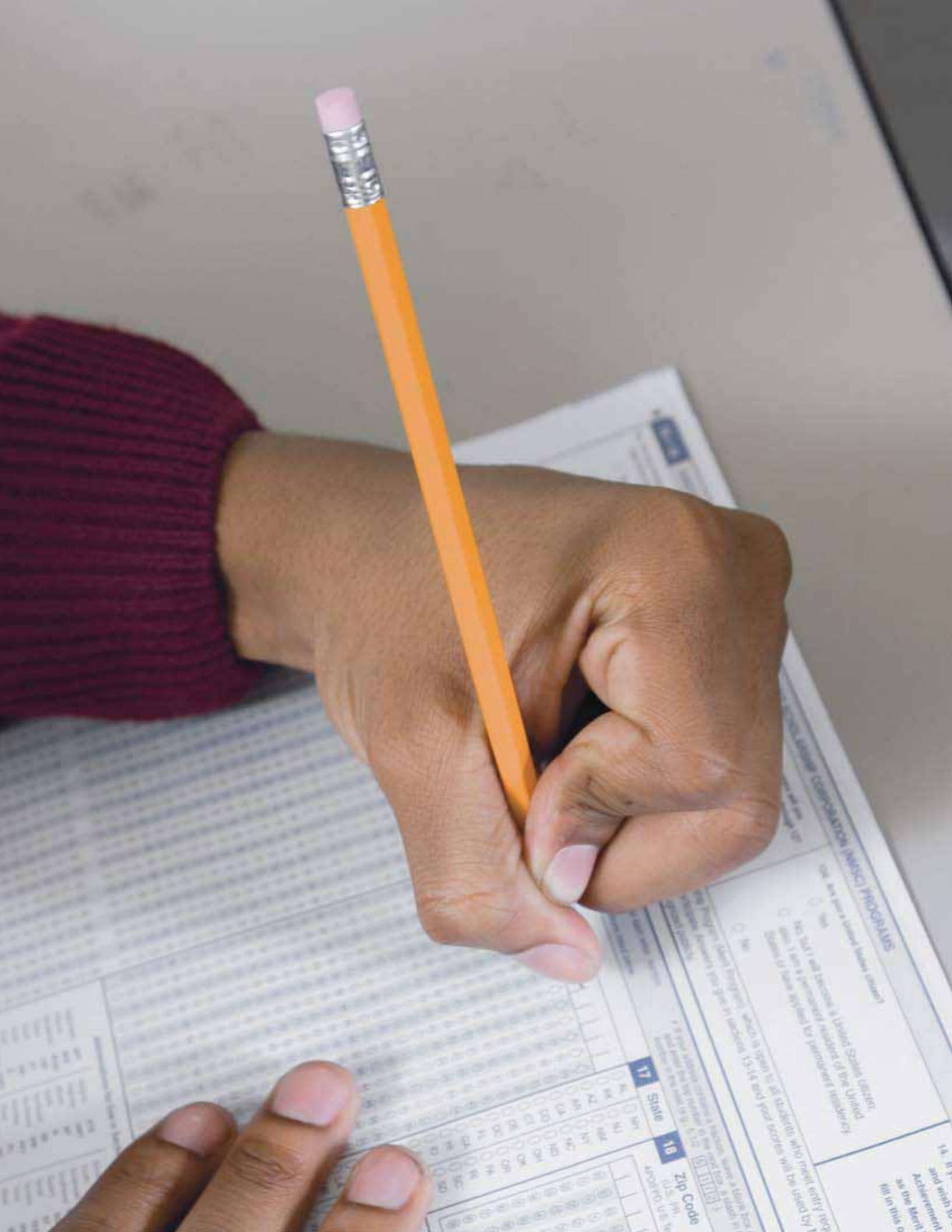
Setting **Local Cut Scores** on the SAT Reasoning Test™ Writing Section

FOR USE IN COLLEGE PLACEMENT AND ADMISSIONS DECISIONS



MARCH 2006

Deanna L. Morgan
Associate Research Scientist
The College Board



CONSIDERATION (INVESTIGATOR) PROGRAMS

Are you a United States citizen, permanent resident, or U.S. Department of Education-approved international student?

Yes

No

Do you have a valid passport?

Yes

No

Do you have a valid visa or other travel document?

Yes

No

17 State AL AK AZ AR CA CO CT DE FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NE NH NJ NM NV NY NC ND OH OK OR PA RI SC SD TN TX UT VT WA WI WY

18 Zip Code (U.S. only)

Abstract

The introduction of the SAT Reasoning Test™ with a writing section in March 2005 and the concomitant elimination of the SAT® Subject Test in Writing after January 2005 have led many colleges and institutions to ask for guidance in using the new SAT Reasoning Test writing section scores for college placement and admissions. Standard-setting methodologies provide one possible set of processes that can be used to identify point(s) on a score scale to divide a group of examinees into categories. Many standard-setting methods have been developed, but only one, the Angoff Method (1971) with Mean Estimation for Essays (Loomis and Bourque 2001), is presented in this document. This document presents many of the decisions needed to conduct a standard-setting session and then provides step-by-step directions to enable colleges or institutions to conduct their own local Angoff-based standard setting for the SAT writing section. Many examples of the data collection documents and supplementary preparation and training materials useful during the standard-setting process, including test items for the March 2005 SAT writing section, are provided. Colleges and institutions are strongly encouraged to use the free Admitted Class Evaluation Service™ (ACES™) offered by the College Board for users of College Board tests to collect information regarding the validity of using specific cut scores on the SAT Reasoning Test writing section for placement or admissions for their institutions' applicants and students.



Contents

Introduction	1
Standard Setting	1
Who Participates in a Standard-Setting Study?	3
The Facilitator	3
Subject Matter Experts (SMEs).....	4
Overview of the Standard-Setting Process	5
The Authoritative Body	5
The Just Minimally Competent Examinee.....	5
Performance-Level Descriptors.....	6
The Angoff Method with Mean Estimation for Essays	10
Training the SMEs on the Process.....	11
Consequence or Impact Data.....	12
To Share or Not to Share Results.....	14
Evaluation and Documentation	14
Setting the Cut Score: Step by Step	15
Logistical Details for Consideration	28
Printed Materials.....	28
Support Staff.....	29
Data Spreadsheet	29
Calculating the Cut Score.....	29
Calculating the Standard Error of Judgment	30
Conclusion	30
References	31

Appendixes	32
Appendix A: Information on the SAT Reasoning Test™ Writing Section.....	32
Appendix B: A Sample Schedule for a Standard Setting Using Angoff with Mean Estimation to Set One Cut Score.....	46
Appendix C: A Sample Schedule for a Standard Setting Using Angoff with Mean Estimation to Set Two Cut Scores.....	47
Appendix D: SAT Essay Prompt and Multiple-Choice Items for Writing—March 2005.....	49
Appendix E: Scoring Guide for the SAT® Essay.....	59
Appendix F: Answer Key for the SAT Reasoning Test Writing Section	60
Appendix G: Training Essay Prompt and Multiple-Choice Items for Writing Section	61
Appendix H: Angoff with Mean Estimation of Essay Ratings Training Record.....	65
Appendix I: Training Evaluation Form.....	66
Appendix J: Angoff with Mean Estimation of Essay Ratings Record.....	67
Appendix K: SAT Writing Composite Score Conversion Table	68
Appendix L: Final Evaluation Form for the SAT Reasoning Test Writing Section	69
Appendix M: Dictionary of Terms.....	70

© 2006 The College Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Board. Admitted Class Evaluation Service, ACES, connect to college success, SAT Preparation Booklet, SAT Preparation Center, SAT Reasoning Test, and The Official SAT Online Course are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Visit the College Board on the Web: www.collegeboard.com.

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

Introduction

In March 2005, the College Board introduced a new writing section to the SAT Reasoning Test. The writing section contains 49 multiple-choice questions and one 25-minute essay. The multiple choice component is broken into two sections, one 25 minutes (35 questions) in length, and the other 10 minutes (14 questions) in length. The College Board recognizes that institutions¹ may be eager to begin using the new writing scores in their admissions and placement decisions, but cautions that until a yearly cohort of data can be collected and analyzed any normative results may not be representative. It is strongly recommended that institutions collect data and after one year perform a validity study through the College Board's free Admitted Class Evaluation Service (ACES) to determine placement scores.

Although a number of institutions have used the SAT Subject Test in Writing as part of their admissions and placement procedures, the SAT Subject Test in Writing was discontinued after January 2005 with the introduction of the new writing section to the SAT Reasoning Test. These institutions, therefore, desire a way to use the SAT Reasoning Test writing section while the first year's data are being collected and analyzed. Institutions are cautioned against using their existing *cut scores*² for the SAT Subject Test in Writing with scores from the new SAT writing section due to a lack of comparability between scores from these two different assessments. The College Board is providing the following guidelines for establishing cut scores through standard setting on the writing section of the SAT Reasoning Test so that institutions have guidance in using writing scores in their admissions and placement procedures before an appropriate validity study can be performed.

This document is organized into two sections. The first section provides general information on standard setting and the rationale behind some of the procedures and practices involved. The second section is a step-by-step guide to the standard-setting process and provides more specific details about exactly how to conduct the standard-setting study.

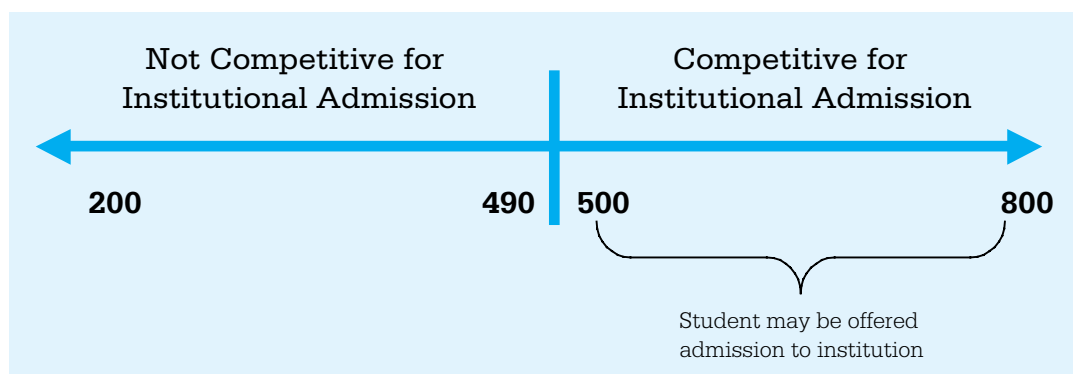
Standard Setting

Standard-setting is the name given to a set of methodologies that may be used to establish a cut score to separate examinees into adjacent groups or categories. Frequently, it is necessary to identify a single point on a test-score scale where examinees below the point are considered to be members of one group and examinees at or above the point are considered to be members of another group. An example of this may be the identification of a test score such that examinees scoring at or above the score are considered competitive for admission to a given institution and examinees scoring below the score are considered not competitive.

1. In this document the word *institution* is used to refer to the broad assortment of higher education entities, e.g., colleges, universities, two-year/four-year colleges, scholarship programs, etc.
2. Italicized words appear in the Dictionary of Terms in Appendix M.

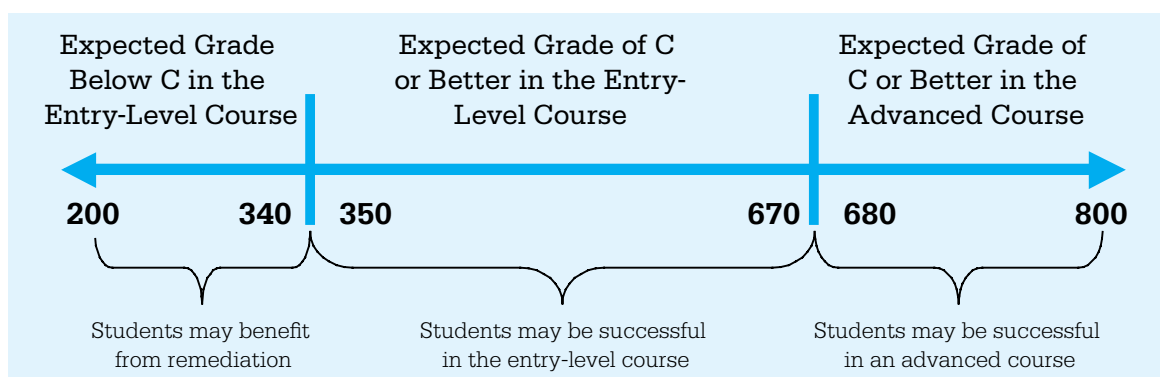
It is important to note that the College Board, in agreement with the *Standards for Educational and Psychological Testing* (APA/AERA/NCME, 1999), recommends that decisions be made based on multiple sources of information and not a single test score. The *Standards* state that “a decision of characterization that will have major impact on a student should not be made on the basis of a single test score.” (p. 146.) However, if the institution decides to use a test with a cut score as one of the multiple sources, then this example may be applicable to admissions decisions and uses only one cut score. (See Figure 1.)

Figure 1. Diagram of a Hypothetical Admissions Decision Using a Cut Score of 500³



Other decisions, such as course placement, may need multiple cut scores. It may be desirable to separate students into multiple groups that are aligned to course placement. An example may be the use of two cut scores to separate students into those who should begin in the entry-level course, those who should begin in a remedial course, and those students who may be successful in a more advanced course. (See Figure 2.)

Figure 2. Diagram of Hypothetical Placement Decisions Using Two Cut Scores⁴



3. The cut score of 500 was arbitrarily chosen for this example and should not be interpreted as a recommendation for the placement of the cut score in an admissions decision.

4. The cut scores of 350 and 680 were arbitrarily chosen for illustrative purposes and should not be interpreted as a recommendation for the location of cut scores for placement purposes.

Many standard setting methods exist to facilitate the identification of a test score to be used as a cut score. Cizek (2001), Morgan and Michaelides (2005), and Hansche (1998) provide procedures and explanations for standard-setting methods beyond the one elaborated upon in this document. Each standard-setting method has both advantages and disadvantages that are specific to its use and methodology that should be carefully considered before choosing a method for use. This document elaborates on the *Angoff Method (1971) with Mean Estimation for Essays* (Loomis and Bourque 2001) due to the relative ease of use, the abundance of research on the method, the large precedence for its use, and the relatively low data requirement for the method.

Who Participates in a Standard-Setting Study?

A *standard-setting study* is one part of a more comprehensive standard-setting process. This part, the study, has two key roles that must be filled for the standard-setting process to proceed with minimum bias in the judgments of the *subject matter experts* (SMEs) and maximum validity for the recommended cut score. The first role is that of an external *facilitator* to oversee the standard-setting process and ensure that appropriate procedures are followed and documentation is maintained. The second role is that of the SMEs who, after sufficient and appropriate training, provide the judgments used to form the recommended cut score. Note, the outcome of a standard-setting study is a recommendation that will be used by the *authoritative body* (described below), which has the responsibility of considering many pieces of information, including the recommendation, to establish the final standard.

The Facilitator

The facilitator is a person with specific skills who is outside the process and does not have an immediate stake in the outcome of the standard setting. The facilitator may be, for example, the director of institutional testing at the institution, a faculty member in the department of educational psychology, or any other individual familiar with the standard-setting process who has the ability to train the SMEs to perform an unfamiliar task, to elicit full participation from each participant, and to ensure that no one participant or set of participants dominates the process. The facilitator must not have an immediate stake in the outcome of the standard setting. The facilitator could be someone from the admissions office; however, this may have the appearance of bias or lead to real bias in the process and results.

The facilitator's role is to ensure that the procedures and tasks that take place during the standard setting are performed as intended to maintain the validity of the process. This includes

not only training the SMEs on the standard-setting tasks they will be performing but also monitoring large- and small-group discussions and providing the SMEs with information that may need to be considered in the standard-setting process. The facilitator should be familiar with standard setting and its requirements and especially familiar with the specific standard-setting method that will be used, in this case the Modified Angoff with Mean Estimation for Essays. In addition to the facilitator not having a stake in the outcome of the standard setting, the facilitator should not be in a position of authority over the SMEs. The standard-setting process is not one with which the SMEs are familiar. It takes some training and involves a question-and-answer period during which the participants need to feel free to express themselves. In addition, the process requires and encourages a large amount of discussion, and the SMEs must feel free to express their opinions. The facilitator should not provide any judgments during the process or interject any personal opinions that may influence the judgments of the SMEs.

Subject Matter Experts (SMEs)

The SMEs should be knowledgeable about the examinee population and the skills and knowledge required of students in relation to the decisions being made. If the purpose of the standard setting is to set a cut score for an examination as part of the procedure to determine eligibility for admission, then the SMEs should be knowledgeable about the skills and knowledge required of students entering college at that institution and should represent faculty from an array of disciplines taught to entering students. However, if the purpose of the standard setting is to determine course placement in a specific content area, then the SMEs should be experts in the content area under consideration.

Ideally, the SMEs will be faculty members of the institution(s) that will use the resulting cut scores. The SMEs should be representative of the college or institution for which the admissions decisions are being made. Representation should be considered in terms of gender, race/ethnicity, tenure (both new staff and veteran staff) and, in cases where the cut score may be intended for multiple campuses or locations, geographical location and campus size. For example, if the cut score will be used systemwide, then representatives from around the system, not just from the main campus, should be included; in addition, representatives from both two-year and four-year campuses are recommended. The more representative the panel of SMEs, the more generalizable and valid the results will be. The panel of SMEs should consist primarily of faculty currently teaching in the subject area. The standard-setting study involves a period of training, and the rest of the process should not be undertaken until all participants have had their questions answered and have established a level of understanding sufficient to perform the duties with confidence and competence. In addition, the process encourages discussions and interactions so the panel should have at least 15 members for the purposes of representation but no more than 30 to allow every panelist to participate and contribute.

Overview of the Standard-Setting Process

The following sections will present information about the process of standard setting. For the purposes of clarity, this document will focus on setting a single cut score; however, it should be noted that if multiple cut scores are desirable, the same process would be used but would need to be repeated for each cut score to be set. Standard setting has a strong tie to policy, and many decisions will need to be considered prior to the start of the standard-setting process. These policy decisions (e.g., how many cut scores are needed, whether to provide impact data, etc.) must be made before beginning the standard-setting process to avoid the risk of making a decision based solely on the outcome of the process and therefore introducing bias into the results. Following the overview, a step-by-step guide is presented to aid users in conducting a standard setting at their institution.

The Authoritative Body

The SMEs convened for the standard setting are content experts and serve to provide a recommendation based primarily on content for the placement of the cut score. The final cut score that will be adopted for use must be approved by someone with the authority to make policy decisions. The authoritative body makes several decisions related to the standard setting but should not participate in the actual standard-setting session. The authoritative body may be a single person but generally is a small group of people. The identity of the authoritative body will differ based on policy and procedure at each institution. Some possibilities for authoritative body members include: the director of admissions, the president of the institution, the board of regents, the vice president of academic affairs, or some combination of those listed. Each institution will have to determine what person or group would best fit their situation. The key point in deciding the identity of the authoritative body is ensuring that the person(s) have the authority and knowledge to make the final decision on the location of the cut score(s).

The Just Minimally Competent Examinee

The objective of the standard-setting process is to identify the point on the score scale that separates examinees who meet the specified qualifications from those who do not. In the case of the SAT writing section, this may be conceptualized as those examinees who exhibit sufficient proficiency in writing to be placed in the entry-level English course and those who do not exhibit sufficient proficiency to be placed in the entry-level English course and therefore may need remediation. This may be shortened to think of those who meet or exceed expectations and those who do not meet expectations. Generally each group (those who meet or exceed expectations and those who do not) contains examinees who obviously belong in one specific group, for

example, very low scorers or very high scorers. However, each group will also contain a number of examinees who either exhibit just enough proficiency to be placed in the entry-level course or who lack the proficiency, but just barely, to keep them from being placed in the entry-level course. Identifying the point on the score scale that signifies just enough proficiency for placement in the entry-level course is the task of the SMEs. Students who exhibit just enough proficiency for placement in the entry-level course are known as *just minimally competent examinees*. When setting a cut score the SMEs should make all decisions with the just minimally competent examinee in mind. Part of the training and discussion entails developing and understanding the definition and description of the just minimally competent examinees, described below.

Placement decisions often require that multiple cut scores be determined. The number of cut scores needed is a policy decision. In the example given earlier in Figure 2, students are being placed into one of three courses: remedial, entry level, or advanced. For this example, two cut scores are needed. The first cut score separates students who are just minimally qualified for the entry-level course from those who do not qualify for the entry-level course and may need remediation. The second cut score separates students who are just minimally qualified for the advanced course from those who do not qualify for the advanced course. The next section discusses the definition of students who belong in each category.

Performance-Level Descriptors

SMEs bring a diverse set of experiences with students and courses, and a variety of opinions into the standard-setting process. While this diversity increases the generalizability of the standard-setting results, it may also introduce a variation in initial definitions of the just minimally competent examinee. Consider, for instance, a group of faculty members teaching the same course at a college or university. The course may use the same curriculum and materials, but it is not uncommon for the requirements necessary to earn a grade of A from one professor to differ slightly, or dramatically, from the requirements necessary to earn a grade of A from another professor. Therefore, it is likely that when asked to think of the just minimally competent examinee, each SME will picture this hypothetical person differently. As a result, it is recommended that prior to the task of standard setting, the SMEs create a set of *performance-level descriptors* that will guide the process and serve to calibrate the SMEs prior to the standard-setting task.

The type of performance-level descriptors needed is determined by the purpose each institution has identified for the cut scores. The facilitator should begin this task by giving a brief introduction to the SMEs on the goal of the standard-setting process, an overview of the test (number of questions of each type, overview of the *scoring guide* used for the essay, etc.), the intended use of the resultant cut score(s), and an overview of the approval/adoption process that will be conducted once the SMEs

have made their recommendation.⁵ Once the SMEs have been informed of their role in the process and the goal of the session, the facilitator should ask the SMEs to verbalize the characteristics in terms of knowledge and skills in writing that a just minimally competent examinee would be expected to exhibit. It is often helpful to conduct this as a brainstorming session by recording all ideas until few additions are forthcoming and then going through the list one idea at a time asking the SMEs to consider how relevant it is to a just minimally competent examinee.

In the case of admissions decisions only one cut score is typically needed and therefore two performance-level descriptors are required. The question for the SMEs to consider in refining their list of characteristics into performance-level descriptors may be: Is the idea absolutely essential for success in (admission to) college, or is it really good to have but not essential?

In the case of placement decisions, multiple cut scores may be needed to separate students into groups for each course level under consideration. Therefore, multiple performance-level descriptors will be necessary. Begin by considering the situation where all entering students are placed in the entry-level course. Some of these students will do very well and may have possessed sufficient knowledge and skills so that they could have taken a more advanced course rather than the entry-level course and still have been successful. Other students will be successful in the entry-level course but probably would not have been successful if placed into a more advanced course instead. The third group is comprised of those students who were unsuccessful in the entry-level course because they did not possess sufficient knowledge and skills to succeed and would have benefited from remediation in a lower-level course. The institution may be interested in identifying the cut scores that will separate students into these three groups to allow differentiated instruction where needed. Developing a set of performance-level descriptors for this situation is more complicated than the essential pass/fail decision needed for an admissions decision.

The creation of performance-level descriptors for placement decisions must begin with a policy decision for how the institution defines success in a course. Typically, success is defined in terms of the grade achieved in the course. For the scenario described above and illustrated in Figure 2, two cut scores will be needed: the cut score representing the just minimally competent student who is successful in the entry-level course and the cut score representing the just minimally competent student who is successful in the advanced course.

How should success be defined in each of these cases? Would a student receiving a grade of C in the entry-level course be considered successful, or does it require a grade of B? The choice of using a grade of C or a grade of B is a policy decision. At some institutions the decision may be to place the cut score using a grade of A or a grade below C to define success. However, typically a grade of B or a grade of C is chosen to define success. Consider the case where the definition of success is a grade of C or better in the entry-level course. This would signify that students earning scores identified to be at the C level or higher would be placed in the entry-level course, and

5. A complete list of the training steps is included in a later section.

students earning scores identified to be below the C level would be placed in the remedial-level course. To help the SMEs clarify the difference in students who are just minimally competent for placement in the entry-level course and those who lack competence for placement in the entry-level course, the SMEs should be asked to develop a performance-level descriptor for each group of students: How do “C” students perform and what do they know versus how do “below C” students perform and what do they know?

The second cut score would also require that a policy decision be made as to the level of success needed to be placed in an advanced course. The level of success should be in terms of performance in the advanced class. Typically this cut score would also be defined as the score that identifies a student who is just minimally competent to earn a grade of C in the advanced course. Institutions may have policy reasons for using a grade other than C to define success in the advanced course. It is important that the decision for the level of success be defined and communicated to the SMEs so that they can create appropriate performance-level descriptors for that level of success.

The set of performance-level descriptors should:

- Describe what students at each level should reasonably know and be able to do.
- Relate directly to the content standards, course prerequisites, and course requirements.
- Distinguish clearly from one level (remedial course) to the next (entry-level course).
- Be written in positive terms.
- Be written in clear and concise language without using nonmeasurable qualifiers such as often, seldom, thorough, frequently, limited, etc.
- Focus on achievement.

Figure 3 provides an example of a set of performance-level descriptors that may be appropriate for use in setting cut scores for the SAT Reasoning Test mathematics section. Mathematics was chosen so as not to influence any institution’s deliberations about the specific characteristics of student writing appropriate at each performance level for their specific situation. Notice that while a lot of similarity exists between the performance-level descriptors at each grade level, the level of complexity or abstraction varies and that an exact one-to-one correspondence between the performance-level descriptors at each level is not necessary. Five statements may be needed to accurately describe the performance at one grade level while four or even six statements may be more appropriate at another grade level. Of primary importance is developing a set of performance-level descriptors that the SMEs are comfortable with and believe they can work with during the standard-setting process. For additional information on writing performance-level descriptors and to see other examples that have been used in standard-setting studies, see Hambleton (2001) or Hansche (1998).

Figure 3. A Mathematics Example of Performance-Level Descriptors for Use in a Standard-Setting Study

The Minimally Competent A-level Student

- They have a strong conceptual understanding and mathematical ability in algebra and geometry.
- They have the ability to solve and model problems in real-world and contextual situations effectively.
- They can synthesize concepts, processes, and procedures to solve complex and nonroutine problems in contextual and real-world settings.
- They are able to make connections between abstractions and concrete situations.
- They are able to use several components of their understanding at the same time. They can break down and keep track of the individual components that build toward the final answer. They will give the completed answer in proper form if there is one.
- They can identify and apply efficient or insightful methods of solution.

The Minimally Competent B-level Student

- They are able to work at an abstract level of understanding.
- They can synthesize concepts, processes, and procedures, but usually within the realm of routine problems.
- They can work with and interpret algebraic and geometric models and have the ability to construct a model for a real-world or contextual situation.
- They have enough conceptual understanding to solve new problems that aren't complex, that is, they can solve new problems that don't have many constituent parts.
- They are likely to make a careless mistake along the way if the problem involves many parts, and they may make a few mistakes when the level of abstraction is increased.
- Although they know how to solve a problem, they do not always apply efficient or insightful methods of solution.

The Minimally Competent C-level Student

- They can, and do, learn the fundamental definitions and theorems, maybe from memorization or by repetition of exercises, but haven't made the connections to the understanding, the analysis of the concept.
- They demonstrate proficiency with simple procedures and algorithms.
- They are memorizers. They can usually only solve routine problems.
- Most modeling by "C" students has probably been shown to them by someone else. They would have difficulty creating their own models.
- They have difficulty with multistep problems because they lack the necessary skills to work through the problem without making a mistake somewhere. A weak background in algebra can cause problems even in those students who have a good understanding of new concepts. Many times students are simply unable to recognize the constituent parts.

The Minimally Competent D-level Student

- Their use of processes is at a concrete numerical skill level.
- They experience difficulty with even simple abstractions, procedures, and algorithms.
- They can only solve problems that they have seen others solve several times before, and they have practiced solving similar problems.
- They do not fully understand the definitions of basic concepts.
- They do not detect errors in their solutions.

Once the performance-level descriptors have been composed, containing only the knowledge or skills that are considered essential for an examinee to be considered just minimally competent, the list either should be posted in a spot visible to all SMEs or copied and distributed to each SME for reference during the standard-setting process. A copy should also be maintained in a folder or binder for reference and documentation in the event that anyone should ever question the institution's use of a specific cut score. The final set of performance-level descriptors provides the meaning in words of the numeric cut score that will be set and adds to the validity of the standard-setting process and the resultant cut score (Hambleton 2001).

After the generation of definitions and training in a large group we suggest breaking into smaller groups of five to seven SMEs for the next steps while the panelists begin making their judgments. This provides an opportunity for cross validation of the recommended cut score between the groups.

The Angoff Method with Mean Estimation for Essays

In the Angoff method, SMEs undergo extensive training (see next section) after which they are asked to picture a just minimally competent examinee (e.g., an examinee on the borderline between two adjacent performance levels) and indicate the probability (probabilities range between zero and 100 percent) that the examinee will correctly answer each multiple-choice test question. Another way to consider this task is to picture 100 just minimally competent students and determine how many of them would answer the question correctly. For the essay, the SME is asked to estimate the average score a just minimally competent examinee would be expected to earn. Each SME must make one judgment for each question. These probabilities are summed for each SME to determine each individual SME's cut score or the *SME Number Correct* (see page 29, step 1, in Calculating the Cut Score). A detailed example of this is provided in the Step-by-Step section of this document. This constitutes the first round of estimation.

Feedback is then provided to the SMEs to provide information for a second round of estimating the probabilities. The *distribution* of the SME Number Correct for each panelist in the small group, without any identifying information, is generally provided to the panel at this time. The SMEs usually exhibit a range of Number Correct, that is, the individual SMEs have recommended a range of suggested cut scores. This information helps panelists understand that there are differences of opinion and maybe differences in interpreting the definitions. The SMEs are given the opportunity for discussion on the feedback provided and then asked to repeat their ratings making any adjustments they would like.

The SMEs are then given feedback in small groups of the *Average Number Correct* (e.g., the current recommendation for the cut score based on only that group's judgments) for the group after the second set of ratings and another opportunity for discussion. Following the small-group discussion, the SMEs are brought back together in the large group for discussion and sharing of main discussion threads from each group.

Afterward, the SMEs are presented with the *impact* data, if the authoritative body of the local institution has decided to provide this information, and a brief opportunity for large group discussion of the impact data. Impact data refers to what the results would be of using the current recommendation(s). That is, impact data is the proportion of students who would be selected, or not, should the cut score be employed. In the case of placement, it would be the proportion of students identified for each class—remedial, entry, and advanced. The facilitator will need frequency distributions of students’ scores on the SAT writing section and must compute the round 2 cut score (see Calculating the Cut Score) to provide impact data.

Then the SMEs have a third and final opportunity to adjust their ratings on each question. When all the third-round judgments have been collected, the SME Number Correct is calculated for each SME based on only the third round of ratings and then averaged across all SMEs to obtain the Average Number Correct for the total group. Because of the inclusion of the essay and the fact that the SAT Reasoning Test is formula scored, an adjustment must be made to the Average Number Correct in order to produce the recommended cut score on the established SAT 200–800 score scale (see Calculating the Cut Score).

Training the SMEs on the Process

The SMEs should be trained on the goals of the standard setting and the intended use of the cut score to be recommended. It is also critical that the SMEs be thoroughly trained and have sufficient practice with the process to be used in the standard setting. The SMEs should be given an overview of the process that includes the tasks they will be performing and the steps that will occur after their recommendation but before the cut score will be considered final. To avoid any potential feelings of having wasted their time or being deceived about the outcome or goal of the process, it is important that the SMEs are aware that their recommendation may be adjusted by an authoritative body prior to implementation. Often a simple explanation of why their recommendation could be adjusted that focuses on the ability of the authoritative body to take into consideration additional information and impact data and to see the larger picture can minimize such feelings or complaints. This is also a good time to ensure that the SMEs know that the materials used in the standard setting and the discussions that take place during the standard setting are considered confidential and that no materials should leave the room. This is less important using the March 2005 SAT writing section enclosed in this document (see Appendix D) due to the fact that it has been fully disclosed, but reasons may still exist that would preclude any information from being divulged about the session.

The training of the SMEs on the process should include an opportunity for them to take the test and to practice using the process. Taking the test gives the SMEs, who are more accustomed to giving exams than taking them, the opportunity to step into the examinees’ shoes and see what it feels like to take the test. Questions can look much less difficult when you are holding the answer

key in your hand. Administering the test to the SMEs familiarizes them with the test questions, makes them aware of any context clues that may affect question difficulty and, without the answer key available, provides them with a less biased view of the true difficulty of the items. Specifically, when taking the test the panelists will pay close attention to every option which tends to sensitize them to the fact that some wrong options will be very attractive to test-takers, particularly the just minimally competent examinee. If possible, the SMEs should take the test under the same time constraints and conditions as the examinees. As the SMEs complete the test, the facilitator should distribute the answer key and allow them to check their own work. Scores should not be collected, but it is a good idea to give the SMEs the opportunity to discuss any questions that they felt were overly difficult or questionable.

The SMEs need the opportunity to practice using the Angoff with Mean Estimation Method prior to performing the task operationally. Often the task may seem simple when explained by the facilitator, but when the SMEs try to implement the process questions can arise that must be addressed prior to the start of the operational standard setting. To provide the practice opportunity, compile a small set of test questions that are similar to those being used in the operational standard setting but not the same questions. A practice set of questions to be used in the training is provided in Appendix G. These questions should be copied so that only one question appears on each page, and the essay prompt, along with the scoring guide, should be the first question in the set. All multiple-choice questions should have the answer clearly marked. Ask the SMEs to provide ratings as instructed in the training for each question using the training record in Appendix H.

When all SMEs have completed the task, ask for three to five volunteers to share their ratings and explain their reasoning. This will allow the facilitator to get an idea of whether the SMEs are providing the ratings as instructed and will also allow the group of SMEs to get an indication of how their ratings may or may not be similar to the group. The practice opportunity and the small debriefing, as the SME volunteers share their ratings, should identify any questions or areas of confusion with the process that need to be resolved prior to the operational standard setting. This is a good time to have the SMEs complete an evaluation form (Appendix I) to allow the facilitator to check on their level of understanding and comfort with the task and to provide another piece of documentation as to the validity of the standard-setting process and resulting cut score (Kane 1994).

Consequence or Impact Data

It is recommended that data be provided to the SMEs on the impact of the cut score that is being considered. Impact data may take several forms and it is a policy decision as to which form it will take if presented. The ability to present impact data depends on the availability of performance data. The simplest form of impact data would present a total number of students applying for

admission and report the percentage of applicants who would be considered to meet expectations for enrollment based on the temporary cut score that was produced in the most recent round of standard setting (typically round 2). For simplicity only, assume that the recommended cut score that came out of the second round of standard setting corresponds to a score of 500 on the SAT writing section. The percentage of students at the institution(s) scoring 500 or higher would be presented to the SMEs along with the corresponding percentage of students scoring below 500 as an indication of the impact of the cut score staying in the current location. This may also be broken down and reported by relevant subgroups, e.g., gender, race/ethnicity, disability, English as a Second Language, etc., to the SMEs. Would the two groups of students, those below 500 and those at or above 500, be similar in terms of subgroup representation? It is a policy decision to determine if the impact data is presented and to determine how much information to present, and for which subgroups.

If the intent of setting a cut score on the SAT writing section is to establish a score to be used in admissions decisions, it is recommended that data be collected for a group of students representative of the institution's student population that contains both the SAT writing section Total Scale Score and the score on an alternate measure of writing, if one is already in use at the institution for each student. This data will allow the SMEs the opportunity to determine if the group of students being included for, or excluded from, admission using the cut score under consideration for the SAT writing will be similar in number and demographics to the group of students who have been included or excluded in the past using an alternate measure of writing proficiency. Although this information is important in helping the SMEs evaluate the validity of using the cut score under consideration for the SAT writing section for their institution, this should be presented to the SMEs only at the end of the standard-setting process.

It should be noted that the use of impact data in standard setting is controversial. Whether or not to use impact data is a policy decision and should be made locally at each institution prior to the start of the standard-setting session. Many experts feel strongly that impact data should be shared with the SMEs to put their ratings into context. Other experts feel strongly that the introduction of the impact data will bias the results and should be avoided (Reckase 2001). While this document will discuss the presentation of impact data for the information of those institutions that may choose to provide impact data, readers should be cautioned that this decision should be made locally by the authoritative body after careful consideration of the issue.

Prior to the provision of the impact data, the standard-setting session has focused on the content of items and how much knowledge of the content a just minimally competent examinee would be able to demonstrate if assessed in the content area. The introduction of impact data to the process serves to quantify the results of the standard-setting process in terms of the effect that the recommended cut score would have on examinees, and it is natural in this context for the SMEs to be concerned if the failure or placement rates look very different from previous

expectations. However, it is important that the SMEs remember all the work that has gone into the identification of a cut score to that point and to consider what any major changes in their ratings would mean in terms of the content knowledge that students would be expected to know and demonstrate. Are the SMEs able to justify accepting a lower level of performance to increase the passing or placement rate? Are the SMEs able to justify expecting a higher level of performance to lower the passing or placement rate? The facilitator is responsible for reminding the SMEs that this is only one more piece of information and should be considered along with the earlier content discussions. The SMEs should also be reminded that they will have the opportunity to revise their ratings one more time before the completion of the standard setting.

To Share or Not to Share Results

At the completion of the standard-setting session, the SMEs have a vested interest in what the outcome of the session will be. During the session, the SMEs were provided feedback as to the location of the recommended cut score. It is common that some of the SMEs will also feel strongly about knowing the location of the final cut score that comes out of the round 3 ratings and will be presented to the authoritative body. Whether or not to share the location of the round 3 cut score is a policy decision. Although it is customary that the materials used and the discussions that occur during a standard setting are considered confidential for reasons of test security, this is less of a concern in this particular standard setting since all materials have previously been released to the public. However, when a recommended cut score is going to be reviewed by an authoritative body before being adopted for use and the possibility exists that the cut score could be modified by the authoritative body, it is sometimes preferable that the SMEs not be aware of the final outcome. This can help to avoid any potential bad feelings on behalf of the SMEs when they know the cut score was modified and by how much.

It is important that the SMEs be made aware of the potential for modification when the cut score goes to the authoritative body, no matter what the decision on whether or not the round 3 cut score will be shared, but it is essential if the round 3 cut score is to be shared. It is also important to inform the SMEs that the authoritative body will have additional information for use in their decision of whether to adopt the cut score as it is or to make modifications, and that the authoritative body will not just arbitrarily decide to change the cut score.

Evaluation and Documentation

The SMEs should be asked to complete a final evaluation form (Appendix L) for the standard-setting session at the conclusion of round 3. The final evaluation provides feedback to the facilitator on how the process may be improved in the future and provides evidence of how the SMEs view the standard-setting session and the resulting cut scores. The evaluation results

should be summarized and included in the materials to be presented to the authoritative body. Strong ratings of confidence in the methods used and the resulting cut score provide support for keeping the recommended cut score as it came out of the standard-setting session. Low ratings of confidence and understanding in the methods provide support for modifying the cut score from the standard-setting session or giving more weight to additional information being considered along with the cut score recommendation. The original evaluation forms should be kept as documentation of what occurred in the event that any legal challenges are made in the future.

Documentation is critical to being able to defend the standard-setting process and the resulting cut score if a legal challenge is made. Documentation should include at least one complete set of all materials and handouts that the SMEs used or had available during the standard-setting session. The names and contact information for the SMEs along with a summary of their demographic characteristics should also be retained for documentation. The completed evaluation and item rating forms should be kept. The completed rating forms can be used to go back and verify that calculations were done correctly and to compare the variation or similarity of each SME's responses. It is good practice to summarize the activities and outcomes of the standard setting in a final technical report as part of the documentation. The technical report should include a thorough description of the procedures and timelines, a copy of the individual group results shared after each round and the large group results shared after round 2, a summary of any impact data provided, a summary of the evaluation form results after the training and at the conclusion of the session, a summary of any information provided to the authoritative body, and the outcome of the authoritative body with appropriate rationale.

Setting the Cut Score: Step by Step

The following sections are intended to provide a step-by-step guide to conducting an Angoff with Mean Estimation for Essays standard-setting method. Where necessary, additional details will be provided, but whenever possible the reader will be referred back to previous areas of this document for specific details. Examples of the agenda for both a one-cut and a two-cut standard-setting session are provided in Appendixes B and C to give an idea of the amount of time that is needed for the standard-setting process.

1. Reserve Facilities/Invite SMEs

It is important that the SMEs have a large space in which they can spread out and be reasonably comfortable for the time needed to complete the standard-setting session. A large room with tables and chairs that can be moved and grouped as needed is ideal. If the cut score will be used systemwide across multiple campuses, arrangements may need to be made for lodging out-of-town

participants. It is also a good idea to make arrangements for snacks or meals to be provided. This is typically the first step in preparing for a standard setting because confirming the availability of suitable space on a particular set of dates must be accomplished to provide sufficient information to the SMEs when trying to obtain a commitment from them to be available.

One of the most important steps in the standard setting is identifying, inviting, and obtaining a commitment from the SMEs. Typically between 15 and 30 SMEs will be needed and, often, as time approaches, attrition will occur. Therefore, it is wise to invite the maximum number of SMEs or to include a few extras, just in case, to ensure that at least 15 are available on the actual dates of the standard setting.

The authoritative body should determine the criteria required to ensure a representative sample of SMEs. Geographic location (e.g., campus), disciplines, teachers of introductory classes or advanced classes, years of experience, gender, and ethnicity are possible criteria. The knowledge that is necessary should be identified. A letter to deans and department heads may be sent soliciting nominations, stating the intentions of the standard setting and listing the knowledge required.

If the cut score will be applicable only to your campus then it may be necessary to include all SMEs on the campus to reach the minimum number. If the cut score will be applicable systemwide or at multiple campuses, then it is suggested that recommendations for representatives from each campus be solicited from the administration at each campus, e.g., the relevant dean of the college for that content area. Soliciting recommendations allows you to indicate on the invitation that each SME was chosen for the honor of participation, and this sometimes helps with obtaining the ultimate commitment from the SME. Be sure to provide ample notice between the invitation and the date of the standard setting, and always ask for a reply confirming acceptance to assist you in determining if additional SMEs need to be invited.

2. Make Policy Decisions

The authoritative body should carefully consider the implications and make final decisions on the following issues prior to the start of the standard-setting session:

- What is the purpose of the standard-setting session?
- How many cut scores will be needed?
- Who will be the facilitator (faculty member, testing director, outside consultant)?
- What qualifications are required of the SMEs?
- What level of proficiency should SMEs keep in mind when creating the performance-level descriptors and setting the cut scores (grade of B, grade of C, other)?

- Will the performance-level descriptors be considered confidential or will SMEs be able to take this information with them when they leave the session?
- Will impact data be presented?
 - How much impact data will be presented and for which subgroups?
 - How will the results of the post-impact data recommendations from the panel be presented to and used by the decision-makers?
- Will the final round 3 cut score be shared with the SMEs?

These decisions should be clearly documented and communicated to the facilitator before the start of the standard-setting session.

3. *Prepare Materials in Advance*

It is important that all materials and data be collected prior to the start of the standard-setting process. The materials include:

- A copy for each SME of the 49 multiple-choice test questions (Appendix D) prepared so that only one question appears on each page, with the set of questions arranged in the order in which they were presented to the examinee and with the correct answer clearly marked. The essay prompt should be accompanied by the scoring guide and appear at the beginning of the set of questions.
- A copy of the test as it was administered to the examinees (this will be a straight copy of the form provided in Appendix D— one question to a page formatting *not* required). The essay prompt should appear first in this set of questions. The answer key in Appendix F should be prepared for distribution to the SMEs upon completion of the test during the training process. Reminder: the SMEs will “take” the test as if they are test-takers, so the key must not be available until this task is completed. The SMEs will “score” their own tests and only they will know the results.
- The formula score to scale score conversion table for the test version that is being used in the standard-setting process (see Appendix K); only one copy is needed.
- The rating form (see Appendix J) will be needed for each SME to record his or her ratings for each question
- The separate set of five multiple-choice writing items and one essay prompt that are supplied in Appendix G. These should be prepared one to a page with the answer clearly marked and the essay placed at the beginning of the set followed by the scoring guide (see Appendix E).
- Impact data, if the decision has been made to provide this information. The data should include frequency distribution of student scores on the SAT writing section.

Compile the materials into tabbed, numbered binders (or other packet of your choice) in the order listed below, giving consideration to the details covered in the Logical Details for Consideration section of this document.

- A. Agenda (See Appendixes B and C for examples)
- B. Information on the purpose of the standard setting
- C. Information on the test (see Appendix A)
- D. Training materials (overheads, handouts)
- E. SAT Reasoning Test Writing Section (see Appendix D)
- F. SAT Reasoning Test Writing Section Answer Key (see Appendix F) Note: the key should not be placed in the binder until after the SMEs take the test during the training.
- G. Training Essay Prompt (with scoring guide) and Questions (see Appendix G and copy one question to a page with the answer clearly marked; see Appendix E for scoring guide)
- H. Ratings Training Record (see Appendix H)
- I. Training Evaluation Form (see Appendix I)
- J. SAT Reasoning Test Writing Section Essay Prompt (with scoring guide) and Multiple-Choice Items (see Appendix D and copy one question to a page with answer clearly marked; see Appendix E for scoring guide)
- K. Ratings Record (see Appendix J)
- L. Final Evaluation Form (see Appendix L)

4. Assign SMEs to Groups

After the SMEs have signed into the standard-setting study, assign them to groups of five to seven people. The standard-setting process will focus on independent input from the individuals, and interaction will occur in both large and small groups. Training and some discussions will occur in the large group, most discussions will occur in the small groups, and the individual ratings of items should occur independently by each SME. Dividing the SMEs into small groups for most discussions serves to provide multiple comparison groups for validation purposes and encourages all SMEs to take part in discussions rather than having one or two SMEs dominate.

Each SME should be assigned an individual ID number to be used to identify all their materials. The ID number should appear on their rating form along with a number identifying the group to which each SME belongs. It is advisable that all the materials the SMEs use in the

standard-setting process also carry their ID number to facilitate ensuring that all materials have been returned. Requiring that the SMEs use the same assigned number on all their materials will also help the facilitator to identify which SMEs have not returned their materials. More tips for handling materials and SMEs can be found in the section Logistical Details for Consideration.

5. Introduction to the Purpose

The standard setting should begin with an introduction to the purpose of the standard setting and a broad overview of the work to be completed during the task. This is also the appropriate time to address any security, confidentiality, and logistical concerns. Representatives from the authoritative body may want to use a few moments of this time to express their gratitude for the participation and expertise of the SMEs. However, any representatives from the authoritative body should leave promptly after offering their gratitude to avoid any undue influence on the process. Representatives from the authoritative body may check in occasionally to see how the session is progressing but should not speak to the SMEs or in any way offer an opinion that may influence the process. Any visitation by the authoritative body should be kept to a minimum.

The introduction should include basic information on the SAT writing section (see Appendix A) and an explanation of how the cut scores will be used. If the use is for placement purposes, then a brief description of the courses involved along with any prerequisites or other information that would be relevant for consideration should be included.

6. Defining the Just Minimally Competent Examinee

The definition of the just minimally competent examinee is one of the first tasks in the standard-setting process. This task sets the stage for the remainder of the tasks in the process and gives the SMEs an initial opportunity to become familiar with each other. Every effort should be made to encourage participation from all SMEs. It is a good strategy to introduce the concept to the large group of SMEs and then ask the small groups to work independently on the brainstorming activity to produce a list of characteristics of the just minimally competent examinee (see the previous sections, The Just Minimally Competent Examinee and Performance-Level Descriptors, on pages 5 and 6). Once the small groups have each developed a list, their lists should be shared with the large group and decisions should be made collaboratively on the final list of characteristics for use in the set of performance-level descriptors. Once a final list is developed it should be readily visible or copied and distributed to the SMEs for reference in the remainder of the standard-setting session.

7. Training the SMEs

Once the performance-level descriptors are finalized, training should begin on the types of items appearing in the SAT writing section and on the specifics of the standard-setting tasks. Begin with the SMEs taking the test under the same time constraints and conditions as those used in

an operational administration (see Training the SMEs on the Process on page 11). As the SMEs finish the test, hand them copies of the answer key to use in checking their work. When everyone is finished, or when the allotted time has concluded, allow the SMEs to discuss any items or item types that they may have found to be overly difficult or confusing. This would be a good time to review the scoring guide for the essay, to remind the SMEs of the section timing used in the operational administration, and to review the penalty for guessing.

When all SMEs are familiar with the test, begin the training on the method. The facilitator should explain the standard-setting task very carefully and be prepared to repeat the explanation often. Refer the SMEs back to the definition of the just minimally competent examinee and remind them that any decisions they make should be made with this group of examinees in mind. The standard-setting task has two parts: 1) providing a rating on the essay, and 2) providing ratings on the multiple-choice items.

The facilitator must be prepared to address the most common concern expressed by SMEs. The SMEs will immediately observe that the standard-setting process is “arbitrary.” In fact, it is a matter of professional judgment that is performed in an objective manner with training, practice, and discussion. The size of the panel reflects the authoritative bodies’ understanding that SMEs will have different experiences and understanding of just minimal competence because of these experiences. Their individual results are important but the collective knowledge is even more so. This is a good time to emphasize how their results will be used by the authoritative body.

The SMEs should begin with the rating of the essay. The SMEs should understand that the profile of each just minimally competent examinee will be different—no two candidates have the same strengths and weaknesses. Therefore, this task requires that the SMEs think of a group of 100 just minimally competent examinees and write the average score that they believe this group of examinees would earn on the essay. The essay rating should be a number between 1.0 and 6.0 and may include one decimal place in the response. For example, 2.1 is a reasonable estimate but 2.15 is not. Remind the SMEs of the time constraints under which the essay is written and to be sure to remind them to take this into account.

The task of rating the multiple-choice items will focus on one item at a time. The SMEs should work independently when providing ratings and move at their own pace. The ratings for the multiple-choice items should also be made with the group of 100 just minimally competent examinees in mind. The SMEs should write the proportion of just minimally competent examinees out of the group of 100 who they would expect to answer each item correctly.

When each task has been explained and all questions from the SMEs addressed, a practice opportunity should be provided. Using the practice set of one essay prompt and five multiple-choice questions, ask the SMEs to independently provide ratings for the essay and each multiple-choice question. When this has been completed, ask for volunteers to share their ratings on

the essay and each question. If discrepancies exist between the ratings, ask the volunteers who assigned the highest and lowest ratings to explain their rationale for the ratings they assigned. The rationales should focus on how the just minimally competent examinee would interact with the item. For example, what skill or lack of skill is related to answering this item in a specific manner? Others in the group may also want to join the discussion. The facilitator should allow enough discussion that any misunderstandings or concerns are addressed and remind the SMEs that this was practice and to please ask any questions they may have about the process. The intention of this task is to have each SME understand the definition of the just minimally competent examinee, the factors that affect the difficulties of the items, and the standard-setting method and process. The facilitator should be careful not to express agreement or disagreement with any of the ratings provided; all facilitator comments should be on the process and should not evaluate the actual ratings. When all questions are answered, the training evaluation form should be distributed (see Appendix I). When completed, the SMEs should hand in the forms, and the facilitator should quickly review the forms to determine if more training on the process is necessary.

8. Round 1 of Ratings

When training is complete, the actual task of assigning ratings that will lead to a cut-score recommendation is begun. The SMEs should have available to them the following materials:

- A copy of the 49 multiple-choice test questions prepared so that only one question appears on each page, with the set of questions arranged in the order in which they were presented to the examinee and with the correct answer clearly marked. The essay prompt should be accompanied by the scoring guide and appear at the beginning of the set of questions for a total of 50 questions in the test packet.
- A ratings record (see Appendix J) to record their ratings for each question.
- The performance-level descriptors of the just minimally competent examinee at each cut score that was developed earlier in the session.

The SMEs are expected to assign the ratings for the essay and each multiple-choice question independently and to work at their own pace. Instruct the SMEs to turn in their rating form once they are finished so that data entry may begin. When finished, the SMEs should either sit quietly while the others finish or leave the room. It is ideal to schedule the session so that lunch, a snack break, or the end of the day will occur at the time the SMEs are expected to finish in order to allow time for data entry of results in preparation for the next round. Data entry typically takes longest for round 1 because all 50 ratings must be entered for each judge. In subsequent rounds, the ratings can be copied and pasted into the spreadsheet for that round and only the items whose ratings were changed will need to be edited.

9. Round 1 Small-Group Discussion

When all the ratings have been entered into a spreadsheet by each SME's ID Number and Group Number (see Table 1), use the steps listed in the Calculating the Cut Score section (see page 29) to calculate the SME Number Correct for each SME. For each small group, separately present a frequency distribution for the members of the group for both the multiple-choice question data (questions 2–50) and the essay data (question 1). (See Figures 4 and 5.) Rating sheets should be returned to each SME at the beginning of the discussion, along with the frequency distribution for the small group. Each group should review the distribution, discuss the high and low ratings within their group, and may choose to compare ratings on individual items along with their rationale for each rating. It is important that the groups be allowed sufficient time to compare results within the group and to discuss the ratings and their perspective as to whether the distribution of SME Number Correct values for their group seems reasonable. No discussion of results should occur between groups at this time. The facilitator should emphasize that the panelists have an opportunity but not a requirement to change their ratings for any item or items if they believe that there is cause to do so. Sometimes a SME did not realize that a *distractor* would be attractive to just minimally competent examinees, for example, and would change the rating downwards. The reverse is possible as well. There should be no effort to lobby a SME to change the responses. If a SME appears to have developed an agenda or has decided to change the definitions of the just minimally competent examinee, the facilitator may want to remind the entire group of SMEs of the definitions.

Round 1 Ratings— Group 1					
	SMEs in Group 1				
Item	SME1	SME2	SME3	SME4	SME5
1 (essay)	3.5	3.0	4.3	2.7	3.8
2	0.35	0.57	0.75	0.97	0.87
3	0.58	0.24	0.68	0.68	0.57
4	0.70	0.36	0.75	0.46	0.68
5	0.80	0.53	0.35	0.57	0.79
6	0.90	0.34	0.85	0.76	0.46
.
.
.
48	0.21	0.87	0.97	0.97	0.89
49	0.74	0.75	0.86	0.86	0.70
50	0.88	0.46	0.79	0.65	0.57
SME Number Correct⁶	25.2	19.3	29.4	31.1	35.7

6. Each of the numbers in this row is the sum of the numbers in the respective column, that is, add all of the numbers in the column.

Figure 4. Example of Small-Group Distribution for Multiple-Choice Data in Table 1

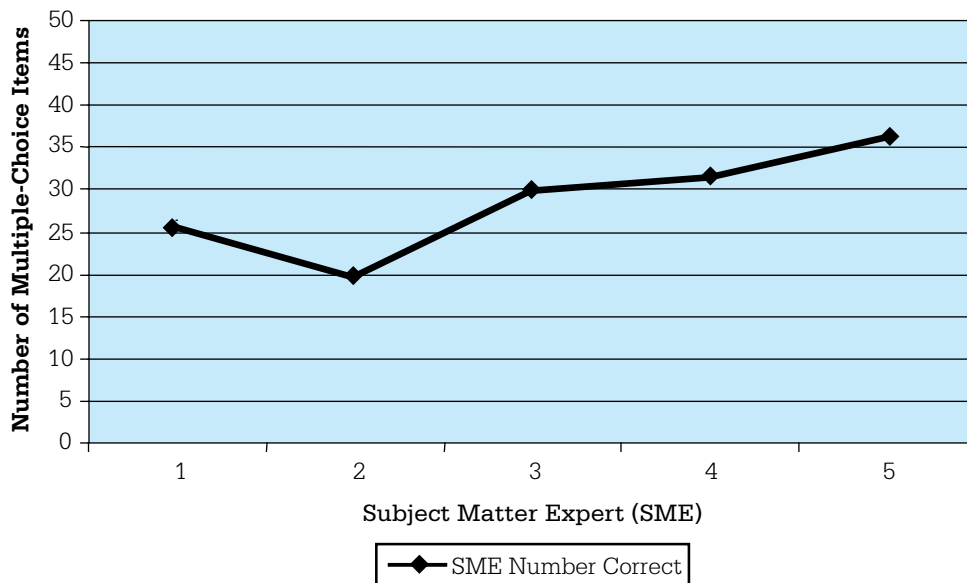
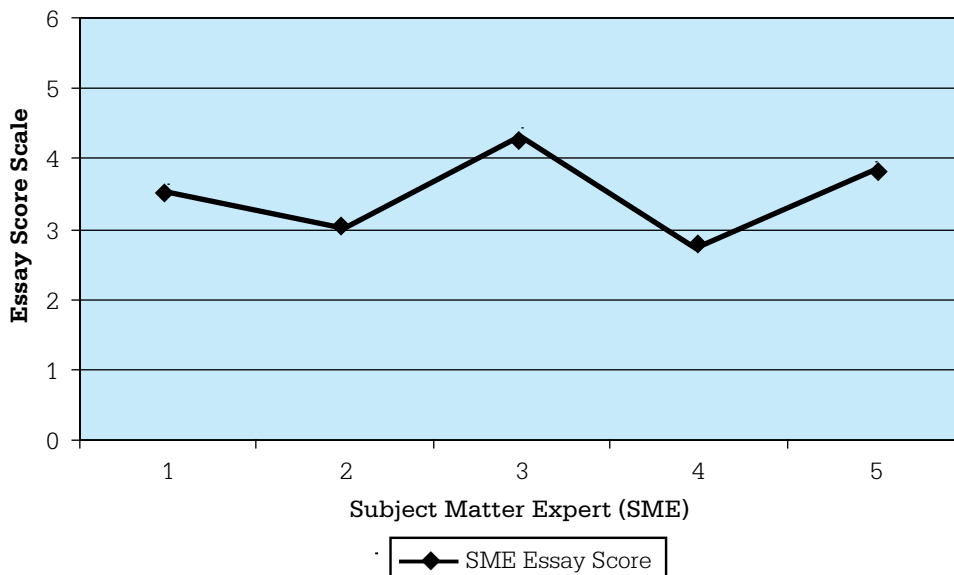


Figure 5. Example of Small-Group Distribution for Essay Data in Table 1

SME Essay Score Distribution for Group 1 in Round 1



10. Round 2 of Ratings

When all groups believe they have had sufficient opportunity to discuss and compare results from round 1, the SMEs are given the opportunity to revise their ratings from the first round.

Typically this second round of ratings will not take as long as the first round, but the SMEs should still assign the ratings independently and work at their own pace. As mentioned above, prior to the start of the round, advise the SMEs that they should feel free to change any and all ratings from round 1, to keep the same ratings they assigned in round 1, or to revise some ratings while keeping others the same. The decision of which item ratings and how many, if any, should be at the sole discretion of each SME. When the ratings are completed, the SMEs should again hand in the rating forms for data entry.

11. Round 2 Small-Group Discussion

When all the ratings have been entered into the spreadsheet for round 2, the Average Number Correct on the multiple-choice questions for each group should be calculated based on the data for round 2 (see Table 2) and reported to each group along with the lowest and highest SME Number Correct in the group. The Average Essay Rating for the group should also be reported along with the lowest and highest essay values in the group. The groups should have sufficient time for discussion within the small group, and each group should be asked to write a short summary of the discussions that have occurred within the group to share with the others in the room during the large-group discussion.

12. Round 2 Large-Group Discussion

At the completion of the small-group discussions for round 2, the facilitator should address the large group and ask that each small group report their temporary cut score and provide a brief summary of the discussions that occurred in the group. This should be the first time that the groups are aware of the ratings of groups outside their own. Keeping discussions in the small groups until just before the third and final round of ratings provides the opportunity to collect multiple cut-score recommendations within the large group to be used as validation of the final recommended cut score.

It is important to remember and to remind the SMEs that there is no RIGHT choice and that the input of all groups is important. Being an outlier group does not mean that the group is WRONG, but it is important that the reason for the discrepancy be shared and discussed in case one or more groups has missed an important perspective that may have resulted in the discrepancy. The facilitator must work hard to make sure that all groups are heard and that no one group dominates the discussion. The facilitator must also be careful not to insert his or her opinion into the discussion.

Table 2. Example of Round 2 Spreadsheet Entry for One Group of SMEs					
Round 2 Ratings—Group 1					
Item	SMEs in Group 1				
	SME1	SME2	SME3	SME4	SME5
1 (essay)	3.5	2.1	3.0	2.7	3.8
2	0.50	0.57	0.75	0.90	0.87
3	0.58	0.24	0.68	0.68	0.70
4	0.70	0.36	0.75	0.46	0.68
5	0.80	0.53	0.35	0.65	0.79
6	0.90	0.34	0.85	0.76	0.60
.
.
.
48	0.45	0.87	0.85	0.97	0.89
49	0.74	0.75	0.86	0.86	0.70
50	0.88	0.46	0.79	0.65	0.70
SME Number Correct	29.1	19.3	31.5	25.3	31.6
Multiple Choice:					
Average Number Correct	27.4				
Lowest Value	19.3				
Highest Value	31.6				
Essay:					
Essay Raw Score	3.0				
Lowest Essay Value	2.1				
Highest Essay Value	3.8				

Note: Bold type indicates values that the SME changed in Round 2.

13. Round 2 Presentation of Impact Data (If applicable)

Prior to the standard setting the decision on whether or not to present impact data should have been made along with the determination of how much data and for which subgroups. The impact data should be presented in reference to the temporary cut score after round 2 that is calculated on all SMEs, not only within small groups. Using the example provided in Table 2 and following the steps as outlined in the section Calculating the Cut Score, the temporary cut score for the previous example would be 460⁷ as shown in Table 3. Depending on the decision made by the authoritative body, if this was a cut score to be used for placement, the impact data may include the number of examinees at the institution who scored 460 or above and would be placed in the entry-level course and the number of examinees at the institution who scored below 460 and would be placed

7. The score of 460 used in the example was calculated using data created for an example only and should *not* be considered a recommendation for the placement of the cut score.

in a remedial course. Additional information by subgroup may include the number of males and females in the 460 and above group and in the below 460 group, or similar information for relevant ethnic or racial subgroups.

Table 3. Example of the Temporary Cut-Score Information Needed to Provide Impact Data to the Large Group					
Round 2—All SMEs for Impact Data Reference					
	All SMEs				
Item	SME 1	SME 2	...	SME 29	SME 30
1 (essay)	3.5	2.1	...	1.7	4.2
2	0.50	0.57	...	0.94	0.96
3	0.58	0.24	...	0.37	0.87
4	0.70	0.36	...	0.45	0.68
5	0.80	0.53	...	0.74	0.74
6	0.90	0.34	...	0.76	0.86
.
.
.
48	0.45	0.87	...	0.65	0.89
49	0.74	0.75	...	0.77	0.70
50	0.88	0.46	...	0.86	0.70
SME Number Correct	29.1	19.3		17.7	34.6
Multiple Choice (MC):					
Average Number Correct	25.2				
Lowest Value	17.7				
Highest Value	34.6				
Essay:					
Essay Raw Score	2.9				
Lowest Essay Value	1.7				
Highest Essay Value	4.2				
Total Number MC items	49				
Average Number Correct	25.2				
<i>Average Number Wrong</i>	23.8				
<i>Formula Score Correction</i>	5.95				
Writing MC Raw Score	19.25				
Writing MC Raw Score (Rounded)	19				
Essay Raw Score (Rounded)	3				
Round 2 Recommended Cut Score	460				

14. Round 3 of Ratings

Following the small- and large-group discussions of round 2 and the presentation of impact data, if any, the SMEs have one additional opportunity to revise their ratings from the previous rounds. Typically, each round of ratings will result in fewer changes so less time is usually needed in round 3. However, it is still important that the SMEs work independently and at their own pace when assigning ratings. The results of round 3 will produce the final cut-score recommendation.

15. Setting Multiple Cut Scores

If multiple cut scores will be needed, then the process must be repeated for each additional cut score. The training will not need to be repeated as long as the same group of SMEs is being used during the same standard-setting session. However, if a different group is used or if the session for the other cut score is to be held at another time, then the training must be repeated. For one continuous standard-setting session where the same SMEs are being used to set multiple cut scores, the performance-level descriptors may all be written at the beginning of the standard setting with a brief review session focusing on the level of the next cut score prior to beginning round 1 or you can choose to create only the performance-level descriptors relevant to the cut score that is being set and then repeat the process to develop the remaining performance-level descriptors when the SMEs begin working on the next performance-level descriptor. If the performance-level descriptors are created by one group of SMEs and it is decided to use the same performance-level descriptors again with another group of SMEs, then the new group of SMEs must at the least have the opportunity to review and edit the performance-level descriptors before they begin providing ratings so that the SME calibration can occur (see Performance-Level Descriptors).

16. Provide Results to the Authoritative Body

The results from the standard-setting session should go to the authoritative body for review and final adoption. The materials for the authoritative body should include, but not be limited to, the following:

- A complete set of the materials used in the standard-setting session.
- The recommended cut score produced after each round with the *Standard Error of Judgment* (SEJ) (see Dictionary of Terms, Appendix M) and the *Standard Error of Measurement* (SEM) (see Appendix M) provided. The scale score SEM for the March 2005 writing section is 40.
- A list of the SMEs and a summary of their demographics.
- A summary of the results of the evaluation forms after the training and at the conclusion of the session.

- A complete copy of the impact data provided to the SMEs along with additional information by subgroups, if that was not part of the original presentation. Even if the decision was made not to provide impact data to the SMEs, it is still preferable to provide this information to the authoritative body.

The authoritative body should have as much information as possible to use in making decisions about the final cut score to be adopted. The cut score recommended by the standard-setting session may be adopted as recommended by the committee. However, a variety of reasons may exist for why modification is necessary. It is not advisable to modify the recommended cut score by more than ± 2 SEM or ± 2 SEJ. When providing the recommended cut score to the authoritative body, the SEM and SEJ should be provided at a minimum. It is useful to compute what the recommended cut score would be at several possible points (as shown in Table 4) and to compile the results for the impact data using each of the resultant potential cut scores as the reference. This facilitates the review and discussion of the cut score and any possible modifications.

	Recommended				
	-2	-1	Cut Score	+1	+2
Standard Error of Measurement (SEM)	380	420	460	500	540
Standard Error of Judgment (SEJ)	360	410	460	510	560

Note: The SEM used in the example is 40 and the SEJ is 50.

Logistical Details for Consideration

Preparation is the key to a successful standard setting. The more information you can prepare in advance the better and more smoothly the standard setting will run. The following section provides basic advice on preparing for the standard-setting session.

Printed Materials

To the extent possible, it is recommended that all printed materials for the SMEs be copied and placed into a large binder in the order in which they will be used. Tabbed dividers may be used to separate the materials into sections, such as Agenda and Introductory Materials, Training Materials, Test Questions and Rating Forms, and Evaluation Forms. The use of a binder saves time during the process by reducing the need to pass out materials at each step. Any material that is considered confidential and should not leave the room may be copied on brightly colored paper to facilitate the identification of the materials. For example, the agenda and any handouts may be on plain white paper but any test questions may be printed on goldenrod paper for easy identification.

At the completion of the standard-setting session, all materials not needed for documentation purposes should be disposed of in a secure manner.

Each binder or packet of materials should be marked with a unique identification number and the SME receiving the packet should be assigned the same unique identification number to assist in tracking materials as they are checked in or out for use. With the exception of the test the SMEs take during training, all test questions should be printed one to a page with the answer clearly marked with an asterisk or other identifying mark beside the correct response.

Support Staff

In addition to the facilitator, it is helpful to have other staff members available during the standard-setting session. Particularly useful for helping the process run smoothly is a data entry person. The data entry person has the responsibility of entering all the item ratings from each SME in a timely manner and producing the summary results after each round for use in the discussion. It is also helpful to have a support person available to make copies and handle any details with catering and checking materials in and out for use in the standard setting and, generally, to oversee any issues that may arise during the course of the standard setting so the facilitator is able to stay focused on the standard setting itself. The support person or the facilitator should have the responsibility of checking the data that has been entered at each round prior to providing feedback to the SMEs to ensure that the feedback is accurate.

Data Spreadsheet

A data spreadsheet is an essential part of any standard-setting session. The spreadsheet serves as a record of the data collected during the session and can be invaluable for allowing the quick computation of results to be provided as feedback to the SMEs during the process. Item ratings should be entered for each SME. Entering the ratings for all SMEs in the same group together facilitates group-level calculations of the mean and standard deviation of ratings for each round. If impact data is being provided, a spreadsheet that contains the impact data summarized by score point will ease the task of providing impact data to the SMEs for the cut score that is being considered.

Calculating the Cut Score

The cut score is calculated from the SMEs ratings using the following method. The same method should be used to compute temporary cut scores for each group following rounds 1 and 2.

1. For each SME, sum the ratings provided for each multiple-choice question to estimate the **SME Number Correct**.

2. Sum the **SME Number Correct** for all members of the group and divide by the number of members in the group to produce the **Average Number Correct**.
3. Subtract the **Average Number Correct** from the total number of multiple-choice questions, in this case, **49**, to get the **Average Number Wrong**.
4. Divide the **Average Number Wrong** by **4** to estimate the **Formula Score Correction**.
5. Subtract the **Formula Score Correction** from the **SME Number Correct** to produce the **Writing Multiple-Choice Raw Score**.
6. Round the **Writing Multiple-Choice Raw Score** to the nearest whole number.
7. Sum the ratings provided for the essay from each SME and divide by the number of SMEs to produce the **Essay Raw Score**. Round the **Essay Raw Score** to the nearest whole number.
8. Use the SAT Writing Composite Score Conversion Table (Appendix K) along with the **Writing Multiple-Choice Raw Score (Rounded)** and the **Essay Raw Score (Rounded)** to locate the Writing Scale Score that corresponds to the recommended cut score.

Calculating the Standard Error of Judgment

The Standard Error of Judgment (SEJ) is produced in the following manner and is an indicator of the variability of the recommendations provided by the panel of SMEs in the standard setting.

1. Compute the standard deviation for the set of round 3 judgments for the SMEs.
2. Divide the standard deviation for the set of round 3 judgments for the SMEs by the square root of the number of SMEs.

Conclusion

Following the procedures outlined in this document will result in a recommended cut score for the SAT writing section. Any process resulting in a cut score should begin by a careful study of the reason a cut score is needed and how that cut score will be used. Preparation is the key for the process to run smoothly and sufficient time should be allowed for the process to occur. Sample schedules are provided in Appendixes B and C.

All cut scores should be validated through the collection of documentation and procedures that are internal to the process and through other means external to the process. A good way to validate the use of a cut score is to take advantage of the free Admitted Class Evaluation Service (ACES) offered by the College Board to users of the College Board tests. Additional information on ACES may be found at www.collegeboard.com/highered/apr/aces/aces.html.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Angoff, W. H. (1971). "Scales, Norms, and Equivalent Scores." In *Educational Measurement*, edited by R. L. Thorndike (2nd ed., pp. 508–600). Washington, DC: American Council on Education.
- Cizek, G. J. (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J., Bunch, M. B., & Koons, H. (2004). "Setting Performance Standards: Contemporary Methods." *Educational Measurement: Issues and Practice* 23(4): 31–50.
- Hambleton, R. K. (2001). "Setting Performance Standards on Educational Assessments and Criteria for Evaluating The Process." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. N. (2000). *Handbook for Setting Standards on Performance Assessments*. Washington, DC: Council of Chief State School Officers.
- Hansche, L. N. (1998). *Handbook for the Development of Performance Standards: Meeting the Requirements of Title I*. Bethesda, MD: U.S. Department of Education, Council of Chief State School Officers (CCSSO). Retrieved May 18, 2005, from: <http://www.ccsso.org/publications/details.cfm?PublicationID=131>.
- Kane, M. (1994). "Validating the Performance Standards Associated with Passing Scores." *Review of Educational Research* 64(3): 425–461.
- Linn, R. L. (1978). "Demands, Cautions, and Suggestions for Setting Standards." *Journal of Educational Measurement* 15: 301–308.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service.
- Loomis, S. C., & Bourque, M. L. (2001). "From Tradition to Innovation: Standard Setting on the National Assessment of Educational Progress." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek. Mahwah, NJ: Lawrence Erlbaum Associates.
- Morgan, D. L., & Michaelides, M. P. (2005). *Setting Cut Scores for College Placement* (College Board Research Report No. 2005-9). New York: The College Board.
- Reckase, M. D. (2001). "Innovative Methods for Helping Standard-Setting Participants to Perform Their Task: The Role of Feedback Regarding Consistency, Accuracy, and Impact." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix A: Information on the SAT Reasoning Test Writing Section

From the 2004–2005 *SAT Preparation Booklet*™

The Writing Section

The new SAT writing section will measure a student's mastery of developing and expressing ideas effectively. It will include both a multiple-choice section and a direct writing measure in the form of an essay. The combination of the multiple-choice items and the essay will provide an assessment of writing that takes into account both the student's ability to develop ideas in a thoughtful, coherent, and cogent essay, and his or her understanding of using the conventions of language to express ideas.

The College Board conducted a survey of high school and college teachers on various questions related to reading and writing curricula. The survey was completed by 2,351 teachers.

- 914 were high school English teachers;
- 814 were college English professors;
- 393 were high school English department chairs; and
- 230 were college professors in other disciplines (including professors of history, political science, psychology, and biology).

The survey investigated how frequently certain reading and writing skills were taught in class, the importance of these skills for students entering their first year of college, the kind and frequency of reading and writing activities done in and out of class, and the level of reading and writing proficiency of college freshmen. The survey revealed the following:

- Persuasive/argumentative writing was rated most important and assigned most often in the classroom in both high school and college.
- High school teachers tended to administer multiple-choice tests more often than did college teachers. Most teachers administered short-answer tests, and almost all teachers administered essay tests.
- Teachers rated almost all of the 14 grammar and usage skills included on the survey high in importance, but survey results indicated that classroom work did not focus heavily on these skills.
- College faculty rated their students' reading and writing skills substantially lower than did high school teachers.
- High school teachers assigned significantly more fiction and poetry reading than did college instructors.

The results of the survey were used by test development committees to set specifications for the new SAT, ensuring that the new test will accurately reflect high school and college curricula and classroom practice.

The multiple-choice questions will test:

- improving sentences (25 questions)
- identifying sentence errors (18 questions)
- improving paragraphs (6 questions)

Multiple-choice writing questions will assess a student's ability to use language that is consistent in tense; to understand parallel structure and subject-verb agreement; to understand how to express ideas logically; and to avoid ambiguous and vague pronouns, excessive wordiness, and

sentence fragments. Students will not be asked to define or use grammatical terms, and spelling and capitalization will not be tested. Here are some examples of the kinds of questions that will appear, followed by brief explanations of the correct answers. Note that students will be provided with detailed directions for each type of question. Multiple-choice questions will count toward about two-thirds of the total writing score.

Note: Calculators may not be on a student’s desk or be used on the writing section of the SAT.

Approaches to the Multiple-Choice Writing Questions

- Read the directions carefully, and then follow them.
- Look at the explanations for each correct answer when using the practice materials in this book. Even if you got the question right, you may learn something from the explanation.
- Eliminate the choices you are sure are wrong when you are not sure of the answer. Make an educated guess from those that remain.

Improving Sentences

This question type measures a student’s ability to:

- recognize and correct faults in usage and sentence structure
- recognize effective sentences that follow the conventions of standard written English

Directions

The following sentences test correctness and effectiveness of expression. Part of each sentence or the entire sentence is underlined; beneath each sentence are five ways of phrasing the underlined material. Choice A repeats the original phrasing; the other four choices are different. If you think the original phrasing produces a better sentence than any of the alternatives, select choice A; if not, select one of the other choices.

In making your selection, follow the requirements of standard written English; that is, pay attention to grammar, choice of words, sentence construction, and punctuation. Your selection should result in the most effective sentence—clear and precise, without awkwardness or ambiguity.

EXAMPLE:

Laura Ingalls Wilder published her first book and she was sixty-five years old then.

- (A) and she was sixty-five years old then
- (B) when she was sixty-five
- (C) at age sixty-five years old
- (D) upon the reaching of sixty-five years
- (E) at the time when she was sixty-five

(A) ● (C) (D) (E)

Answering Improving Sentences Questions

Look carefully at the underlined portion of the sentence because it may have to be revised. Keep in mind that the rest of the sentence stays the same. Follow the two outlined steps to answer each Improving Sentences question.

Step 1: Read the entire sentence carefully but quickly and ask yourself whether the underlined portion is correct or whether it needs to be revised.

In the example above, connecting the two ideas “**Laura Ingalls Wilder published her first book**” and “**she was sixty-five years old then**” with the word “**and**” indicates that the two ideas are equally important. The word “**and**” should be replaced to establish the relationship between the two ideas.

Step 2: Read choices (A) through (E), replacing the underlined part with each answer choice to determine which revision results in a sentence that is clear and precise and meets the requirements of standard written English.

Remember that choice (A) is the same as the underlined portion. Even if you think that the underline does not require correction and choice (A) is the correct answer, it is a good idea to read each choice quickly to make sure.

- In (A), the word “**and**” indicates that the two ideas it connects are equally important. No.
- In (B), replacing the word “**and**” with “**when**” clearly expresses the information that the sentence is intended to convey by relating Laura Ingalls Wilder’s age to her achievement. Yes, but continue to look at the other revisions.
- In (C), using the word “**at**” results in a phrase that is not idiomatic. No.
- In (D), the phrase “**upon the reaching of**” also results in a phrase that is not idiomatic. No.
- In (E), the phrase “**at the time when she was sixty-five**” is awkward and wordy. No.

Correct answer: B

Sample Questions

1. Scenes from the everyday lives of African Americans, which are realistically depicted in the paintings of Henry Ossawa Tanner.
(A) Scenes from the everyday lives of African Americans, which are realistically depicted in the paintings of Henry Ossawa Tanner.
(B) Scenes from the everyday lives of African Americans being realistically depicted in the paintings of Henry Ossawa Tanner.
(C) The paintings of Henry Ossawa Tanner realistically depict scenes from the everyday lives of African Americans.
(D) Henry Ossawa Tanner, in his realistic paintings, depicting scenes from the everyday lives of African Americans.
(E) Henry Ossawa Tanner, whose paintings realistically depict scenes from the everyday lives of African Americans.

Explanation

For a sentence to be grammatically complete, it must include both a subject and a main verb. When a sentence lacks either a subject or a main verb, the result is a *sentence fragment*. In this example all options but (C) are sentence fragments.

- In (A), the phrase “**Scenes...Americans**” is modified by the dependent clause “**which... Tanner;**” but there is no main verb.
- In (B), the phrase “**Scenes...Tanner**” contains no main verb.
- In (D), the noun “**Henry Ossawa Tanner**” is modified by “**depicting**” but is not combined with a main verb.
- In (E), the noun “**Henry Ossawa Tanner**” is modified by the dependent clause “**whose... Americans**” but not combined with a main verb.
- (C) is correct. It is the only choice in which a subject “**The paintings of Henry Ossawa Tanner**” is combined with a verb “**depict**” to express a complete thought.

Correct answer: C

2. Looking up from the base of the mountain, the trail seemed more treacherous than it really was.
(A) Looking up
(B) While looking up
(C) By looking up
(D) Viewing
(E) Viewed

Explanation

When a modifying phrase begins a sentence, it must logically modify the sentence’s subject; otherwise, it is a *dangling modifier*. In this example, every option except (E) is a dangling modifier.

- In (A), the phrase “Looking up from the base of the mountain” does not logically modify the subject “the trail.” A person might stand

at the base of a mountain and look up at a trail, but it is illogical to suggest that a trail looks up from the base of a mountain.

- (B), (C), and (D) are simply variations of the error found in (A). Each results in a sentence that illogically suggests that a trail was looking up from the base of a mountain.
- (E) is correct. Although a trail cannot itself look up from the base of a mountain, a trail can *be viewed* by someone looking up from the base of a mountain, so the phrase “Viewed from the base of the mountain” logically modifies the subject “the trail.”

Correct answer: E

Identifying Sentence Errors

This question type measures a student’s ability to:

- recognize faults in usage
- recognize effective sentences that follow the conventions of standard written English

Directions

The following sentences test your ability to recognize grammar and usage errors. Each sentence contains either a single error or no error at all. No sentence contains more than one error. The error, if there is one, is underlined and lettered. If the sentence contains an error, select the one underlined part that must be changed to make the sentence correct. If the sentence is correct, select choice E. In choosing answers, follow the requirements of standard written English.

EXAMPLE:

The other delegates and him immediately
A B C
accepted the resolution drafted by the
D
neutral states. No error
E

(A) (B) (C) (D) (E)

Answering Identifying Sentence Errors Questions

Ask yourself if any of the underlined words and phrases in the sentence contains a grammar or usage error. Follow the two outlined steps in answering each Identifying Sentence Errors question.

Step 1: Read the entire sentence carefully but quickly, paying attention to underlined choices (A) through (D).

- In the example above, “**The other delegates and him**” are the people who “**immediately accepted the resolution**,” and the phrase “**drafted by the neutral states**” describes “**the resolution**.” Check each underlined word or phrase for correctness.
- The phrase “**The other**” correctly modifies the word “**delegates**.”
- The pronoun “**him**” is in the wrong case. (One would not say “him immediately accepted.”) “**Him**” is an error, but go on to check the other choices, especially if you are not sure.
- The word “**immediately**,” which modifies the verb “**accepted**,” is correct.
- The phrase “**drafted by**” correctly expresses the action of the “**neutral states**.”

Step 2: Select the underlined word or phrase that needs to be changed to make the sentence correct.

- Mark (E) **No error** if you believe that the sentence is correct as written. In this case, mark (B) on your answer sheet because the underlined word “**him**” must be changed to “**he**” to make the sentence correct.

Correct answer: B

Keep in mind that some sentences do not contain an error.

Sample Questions

3. The students have discovered that they can
 A B
address issues more effectively through
 C
letter-writing campaigns and not through
 D
public demonstrations. No error
 E

Explanation

- The error in this sentence occurs at (D). When a comparison is introduced by the adverb “more,” as in “**more effectively**,” the second part of the comparison must be introduced by the conjunction “**than**” rather than “**and not**.”
- The other options contain no errors. In (A), the plural verb “**have discovered**” agrees with the plural subject “**students**.” In (B), the plural pronoun “**they**” correctly refers to the plural noun “**students**.” In (C), the preposition “**through**” appropriately expresses the means by which issues are addressed.

Correct answer: D

The sentence may be corrected as follows: The students have discovered that they can address issues more effectively through letter-writing campaigns than through public demonstrations.

4. After hours of futile debate, the committee has
 A
decided to postpone further discussion
 B
of the resolution until their next meeting.
 C D
No error
 E

Explanation

- The error occurs at (D). A pronoun must agree in number (singular or plural) with the noun to which it refers. Here, the plural pronoun “**their**” incorrectly refers to the singular noun “**committee**.”
- The other options contain no errors. In (A), the preposition “**After**” appropriately introduces a phrase that indicates when the committee made its decision. In (B), “**to postpone**” is the verb form needed to complete the description of the committee’s decision. In (C), the prepositional phrase “**of the resolution**” appropriately specifies the subject of the postponed discussion.

Correct answer: D

The sentence may be corrected as follows: After hours of futile debate, the committee has decided to postpone further discussion of the resolution until its next meeting.

Improving Paragraphs

This type of question measures a student’s ability to:

- edit and revise sentences in the context of a paragraph or entire essay
- organize and develop paragraphs in a coherent and logical manner
- apply the conventions of standard written English

Answering Improving Paragraphs Questions

To answer the Improving Paragraph questions that accompany the draft essay, students will need to note what sentences need to be corrected and to know how each of the sentences relates to one another and to the essay as a whole. Follow the outlined steps to answer the questions.

Step 1: Read the entire essay quickly to determine its overall meaning. The essay is intended as a draft, so there will be errors.

Step 2: In answering each question, make sure that the answer about a particular sentence or group of sentences makes sense in the context of the passage as a whole. Choose the best answer from among the choices given, even if you can imagine another correct response.

Directions

Directions: The following passage is an early draft of an essay. Some parts of the passage need to be rewritten.

Read the passage and select the best answers for the questions that follow. Some questions are about particular sentences or parts of sentences and ask you to improve sentence structure or word choice. Other questions ask you to consider organization and development. In choosing answers, follow the requirements of standard written English.

Sample Questions

Questions 5–7 are based on the following essay:

(1) Many times art history courses focus on the great “masters,” ignoring those women who should have achieved fame. (2) Often women artists like Mary Cassatt have worked in the shadows of their male contemporaries. (3) They have rarely received much attention during their lifetimes.

(4) My art teacher has tried to make up for it by teaching us about women artists and their work. (5) Recently she came to class very excited; she had just read about a little-known artist named Annie Johnson, a high school teacher who had lived all of her life in New Haven, Connecticut. (6) Johnson never sold a painting, and her obituary in 1937 did not even mention her many paintings. (7) Thanks to Bruce Blanchard, a Connecticut businessman who bought some of her watercolors at an estate sale. (8) Johnson is finally starting to get the attention that she deserved more than one hundred years ago.

(9) Blanchard now owns a private collection of hundreds of Johnson’s works—watercolors, charcoal sketches, and pen-and-ink drawings.

(10) There are portraits and there are landscapes. (11) The thing that makes her work stand out are the portraits. (12) My teacher described them as “unsentimental.”

(13) They do not idealize characters.

(14) Characters are presented almost photographically. (15) Many of the people in the pictures had an isolated, haunted look.

(16) My teacher said that isolation symbolizes Johnson’s life as an artist.

5. In context, which is the best revision to the underlined portion of sentence 3 (reproduced below)?

They have rarely received much attention during their lifetimes.

- (A) In fact, they had
- (B) Too bad these artists have
- (C) As a result, these women have
- (D) In spite of this, women artists
- (E) Often it is the case that the former have

Explanation

Although sentence 3 is not grammatically wrong, its relationship to the preceding sentence needs to be made clearer. A transitional phrase should be added to emphasize the cause-and-effect relationship between the stated facts—women artists received little attention *as a consequence of* having worked in the shadows of their male contemporaries—and the ambiguous pronoun “**They**” should be replaced with a word or phrase that clearly refers to the “**women artists**” and not the “**male contemporaries**” mentioned in sentence 2.

- (A), (B), and (D) are unsatisfactory because in each case the transitional phrase (“**In fact**,” “**Too bad**,” or “**In spite of this**”) fails to indicate the cause-and-effect relationship. Moreover, both (A) and (B) leave the ambiguity of the pronoun unresolved.

- (E) is unsatisfactory not only because it fails to signal the cause-and-effect relationship but also because it is wordy and illogically combines the adverbs “Often” and “rarely.”
- (C) is correct. The transitional phrase “As a result” clearly indicates a cause-and-effect relationship, and “these women” properly resolves the ambiguity of the pronoun “They.”

Correct answer: C

6. In context, which of the following revisions to sentence 7 is most needed?
- (A) Delete “Thanks to”.
 - (B) Move “Thanks to Bruce Blanchard” to the end of sentence 7.
 - (C) Delete “who.”
 - (D) Change “her” to “Johnson’s.”
 - (E) Change the period to a comma and combine sentence 7 with sentence 8.

Explanation

Sentence 7 is a sentence fragment, with neither a subject nor a main verb to finish the thought it has begun. It says “**Thanks to Bruce Blanchard,**” but it does not say *what happened* thanks to Bruce Blanchard. It should therefore be joined to an independent clause, complete with subject and verb, that indicates what happened as a result of Blanchard’s action.

- (A), (B), and (D) are unsatisfactory because each fails to provide the main verb needed to complete the sentence. Each results in another sentence fragment.
- Although (C) results in a complete sentence, the sentence makes little sense in the context of the paragraph because it suggests that Bruce Blanchard is someone other than the Connecticut businessman who bought the watercolors.

- (E) is correct. This change results in a grammatically complete sentence that indicates what happened thanks to Bruce Blanchard’s efforts: Johnson began to get the attention she deserved.

Correct answer: E

7. In context, which of the following is the best version of sentence 10 (reproduced below)?
- There are portraits and there are landscapes.*
- (A) (As it is now)
 - (B) You can see both portraits and landscapes.
 - (C) Therefore, both portraits and landscapes are among her works.
 - (D) Johnson painted both portraits and landscapes.
 - (E) Among them Johnson has portraits and landscapes.

Explanation

In addition to being vague, sentence 10 contains no noun to which the pronoun “her” in sentence 11 may refer. It should be revised so that Johnson is clearly identified as the painter of the portraits and landscapes.

- (A), (B), and (C) are unsatisfactory because each omits any mention of Johnson.
- Though (E) does mention Johnson, it is misleading in that the words “**Johnson has**” suggest that Johnson is the owner rather than the painter of the portraits and landscapes.
- (D) is correct because it properly identifies Johnson as the painter of the artworks and thus provides an antecedent for the pronoun “her” in sentence 11.

Correct answer: D

The Essay

The essay will assess students' ability to think critically and to write effectively under time constraints similar to those they will encounter on essay examinations in college courses. Students will be given twenty-five minutes to handwrite their essay. The essay portion of the writing section was created with an understanding that an essay written in a short amount of time will not be polished but represents the initial phase of the writing process: the first draft. The essay will count toward roughly one-third of the total writing score.

The essay component asks students to write in response to an essay prompt that is carefully selected so they can respond quickly in a variety of ways. Prompts are easily accessible to the general test-taking population, including those for whom English is a second language. Prompts are free of figurative, technical, or specific literary references; they do not draw on specialized knowledge.

The prompt stimulates critical thinking and will be relevant to any number of fields and interests. It gives students the opportunity to draw on a broad range of experiences, learning, and ideas to support their points of view on the issue in question. Students may write about literature, the arts, sports, politics, technology and science, history, current events, or personal observations, among other topics. Students may accept or reject the idea presented in the prompt to whatever extent they see fit. They may draw on the rhetorical approach that best suits their writing style and purpose. For instance, some students may use an expository or argumentative style; others may structure essays through comparison or contrast, or other techniques. The essay will measure a student's ability to:

- develop a point of view on an issue presented in an excerpt
- use reasoning and evidence based on his or her reading, studies, experience, and observations to support that point of view
- follow the conventions of standard written English

Scoring the Essay

Essays will be scored in a manner that is fair and consistent using a holistic approach. In holistic scoring, a piece of writing is considered as a total work, the whole of which is greater than the sum of its parts. The essay will be scored by qualified readers who will take into account such aspects as complexity of thought, substantiality of development, and facility with language. A reader does not judge a work based on its separate traits, but rather on the total impression it creates. Holistic scoring recognizes that the real merit of a piece of writing cannot be determined by merely adding together the values assigned to such separate factors as word choice, organization, use of evidence, and adherence to the conventions of written English. It is how these separate factors blend into and become the whole that is important. Holistic scoring evaluates this whole equitably and reliably. Readers are trained to be mindful of the conditions under which students wrote the essays. The essay is viewed as a timed first draft and is evaluated as such.

Readers will use the Scoring Guide (see Appendix E, page 59) in conjunction with the sample essays selected for training. The Scoring Guide provides a consistent and coherent framework for differentiating between score points.

Each essay will be scored independently by two qualified readers and will be scored on a scale of 1 to 6 by each reader, with the combined score for both readers ranging from 2 to 12. If the two readers' scores differ by more than one point, the scoring leader will resolve the difference. Essays not written on the essay assignment will receive a score of zero.

Approaches to the Essay

- *Read the prompt carefully, and make sure you write on the topic given.* Essays not on topic will receive a zero.
- *Decide your viewpoint on the topic.* If you have trouble focusing on the main point, try completing this sentence: “When people finish reading my essay, I want them to understand that...” The words you use to fill in that blank might well become part of your thesis statement.
- *Spend five minutes on planning.* Use your test booklet to create a quick sentence outline. Begin with the thesis from the approach above. You have only twenty-five minutes to write your essay, so don’t spend too much time outlining. A little planning time, however, may be essential to make sure that your essay does not wander off topic or stray from your focused thesis statement.
- *Vary the sentence structure in your writing.* Good writing uses a variety of sentence types to make the writing more interesting while showing the relationships between ideas. To give your prose a *mature* character, *vary* the sentence structure that communicates how your ideas are related to each other.
- *Use clear, precise, and appropriate vocabulary.* Appropriate words are accurate and specific, not necessarily long and obscure.
- *Leave time to review what you’ve written.* Although you won’t have time for full-scale revision, do leave a few minutes for rereading your essay and making minor changes in the wording or even in the structure of what you have written.



Students can see additional sample essays online in the SAT Preparation Center™. Visit www.collegeboard.com/srp.

Directions

The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

Your essay must be written on the lines provided on your answer sheet—you will receive no other paper on which to write. You will have enough space if you write on every line, avoid wide margins, and keep your handwriting to a reasonable size. Remember that people who are not familiar with your handwriting will read what you write. Try to write or print so that what you are writing is legible to those readers.

You have twenty-five minutes to write an essay on the topic assigned below. **DO NOT WRITE ON ANOTHER TOPIC. AN OFF-TOPIC ESSAY WILL RECEIVE A SCORE OF ZERO.**

Think carefully about the issue presented in the following excerpt and the assignment below:

A sense of happiness and fulfillment, not personal gain, is the best motivation and reward for one’s achievements. Expecting a reward of wealth or recognition for achieving a goal can lead to disappointment and frustration. If we want to be happy in what we do in life, we should not seek achievement for the sake of winning wealth and fame. The personal satisfaction of a job well done is its own reward.

Assignment: Are people motivated to achieve by personal satisfaction rather than by money or fame? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

Sample Essays

Essay #1: This essay received a score of 6

Even though we live in a capitalist society, I still cannot help but believe, despite my own cynicism, that people are more motivated to achieve something for personal satisfaction rather than monetary gains. Look at Chekov's short story, "The Bet." A man agrees to sacrifice fifteen years of his life in prison in exchange for a million dollars. Obviously his motivation for such an extreme bet is wealth, but by the end of the prison sentence, the man could care less about the money. After years of introspection, of reading Shakespeare, *The Bible*, and textbooks, the man actually comes to despise the money he once sought; the money he signed away fifteen years of his life for. He does not collect his money from the banker, he runs away to be on his own and continue to live the life of solitude he has learned to love, free of money and possessions.

Also, in a psychology class, one of the first things students study when they come to the topic of motivation, is external stimulus versus personal drive. Any textbook will tell one that studies show that a child is more likely to put as much energy as possible into completing a task when it is something that makes him happy, than if he was doing it for a physical reward. A child is more likely to get good grades, if it makes him feel good about himself, than if his parents offer to pay him every time he makes the honor roll. I agree with this theory on motivation because I see it play out everyday in my life. If my older sister had been concerned with money and fame, which reality television tells us every night is important, she would have gone to college after graduating high school. She knew though, that school and learning did not make her happy, and she was not going to suffer through four more years of school just because a college degree could lead to a more successful job. Right now she does not make as much money at her job, but she likes her life and the way she lives; she has more fun answering phones and dealing with other people at work

than she would behind a desk in a classroom. This past year I myself have been forced to look at my priorities as well. I have worked hard in school all my life and have made honor roll semester after semester, because I enjoy it. I have not filled up my schedule with classes I did not want because calculus and economics look good on a college transcript. I had a high enough GPA to join the National Honor Society, but I chose not to join because even though it might have impressed some admissions officers, it was not something that was going to make me happy. Instead I spend my time studying Creative Writing, Art History, and the other subjects I feel truly passionate about.

There is a pleasure principle in psychology, which basically means that one will do whatever will make them most happy or least unhappy. I think that is true, and I feel that the happiness most people seek out is not about money or luxury. Maybe it looks like that from the media, because advertising says that people want to be like Donald Trump, but that is not real life. Real life is my next door neighbor who gardens as a second job for small fees because he loves to be outside, working with his hands in the nice weather. I am sure no one would mind winning the lottery, but to say that it is our primary motivator in life is sad and untrue. A person who is happy and making minimum wage is likely to live longer than someone who spends his or her life working sixty four hour weeks at a stressful job to make money hand over fist. Are some people very driven by money? Yes. Is that more important than the personal satisfaction that comes from doing something good? Literature, psychology, and our personal lives tell us no, and I hope it stays that way.



Students can receive an automated score for their responses to this essay question with The Official SAT Online Course™ at www.collegeboard.com/satonlinecourse.

Why Essay #1 Received a Score of 6

This outstanding essay insightfully and effectively develops the point of view that, “Even though we live in a capitalist society, I still cannot help but believe...that people are more motivated to achieve something for personal satisfaction rather than monetary gains.” The writer demonstrates outstanding critical thinking by focusing on clearly appropriate examples from “literature, psychology, and our personal lives” to support this position. The essay begins by describing Chekov’s “The Bet” as a short story that, through its main character’s changed priorities after “years of introspection” in prison, ultimately places higher value on a “life of solitude...free of money and possessions” than on wealth. The writer continues to demonstrate critical thinking by offering as evidence the psychological example of “external stimulus versus personal drive” and several examples of how this principle has been borne out in a sister’s and the writer’s own lives. The essay concludes by once again drawing on psychology, this time the concept of the “pleasure principle,” as well as an additional example of the writer’s neighbor, to reinforce the idea that “the happiness most people seek out is not about money or luxury” but in “the personal satisfaction that comes from doing something good.” This well-organized and clearly focused essay demonstrates coherence and progression of ideas. The essay consistently exhibits skillful use of language and demonstrates meaningful variety in sentence structure (“After years of introspection, of reading Shakespeare, *The Bible*, and textbooks, the man actually comes to despise the money he once sought; the money he signed away fifteen years of his life for”). Thus, this essay demonstrates clear and consistent mastery and is scored 6.

Essay #2: This essay received a score of 5

I believe that personal satisfaction, through giving, and providing your best effort, has a greater and more rewarding outcome over gaining wealth or fame. This is evident through many public examples of people who are known to our society, as “celebrities” stooping to a standard of drug reliance and materialism. It seems to me that people who have achieved the money and achieved the fame are never quite satisfied with themselves or the situation they are in. Although many people are based on financial stature, there are also many groups or individuals that are based on goals more helpful to our society as a whole.

People who are involved in an organization such as the make-a-wish foundation and Habitat for Humanity are working towards a common goal of giving others opportunities, and changes they wouldn’t normally have. These foundations are non-profit, and work with and around our communities to help advance and improve the lives of others. People who do these projects and strive for the betterment of others are not doing it for themselves, for the money, or for the recognition. They are doing it whole heartedly, without complaint.

To me, people who don’t think of what benefits or rewards they will reap are truly the ones obtaining the most righteous self-achievement. Through their actions and beliefs, I undoubtedly believe that these people are gaining the happiness and fulfillment that even the richest or most famous people wish they had. When you go into a project or any type of action not expecting anything back, the things you do will be paid back to you ten-fold in the grand scheme of things. Doing that action, or job, or project, whatever it may be, to the best of your ability will give you a sense of accomplishment like no other because of the amount of effort and giving you put forth. It is too bad people or groups like this aren’t recognized as much as a person who wins a game show or is in a movie.

Why Essay #2 Received a Score of 5

This focused essay demonstrates strong critical thinking and effectively develops its point of view (“I believe that personal satisfaction, through giving, and providing your best effort, has a greater and more rewarding outcome over gaining wealth or fame”) by offering a well-organized progression of ideas detailing the motivations of and benefits for groups such as “the make-a-wish foundation and Habitat for Humanity,” who “help advance and improve the lives of others.” Additional focus is provided at the essay’s opening and closing by comparing these groups to “many public examples of... ‘celebrities’ stooping to a standard of drug reliance and materialism” because they “are never quite satisfied with themselves or the situation they are in.” However, some lapses in the specificity of support prevent the response from earning a higher score. Therefore, to merit a score of 6, this essay needs to provide additional detailed evidence to more evenly and insightfully develop the point of view. The response exhibits facility in the use of language and variety in sentence structure (“People who do these projects and strive for the betterment of others are not doing it for themselves, for the money, or for the recognition. They are doing it whole heartedly, without complaint”). Overall, this response exhibits reasonably consistent mastery and receives a score of 5.



Students can receive an automated score for their responses to this essay question with The Official SAT Online Course at www.collegeboard.com/satonlinecourse.

Essay #3: This essay received a score of 4

Some people in the world today feel that they need to achieve a goal in order to receive fame and money. While others achieve a goal for their own satisfaction. I personally feel that you should achieve a goal for your own satisfaction and not for that of fame.

The most important thing about someone’s life is to have a positive self-esteem. People need to not worry so much about what others think and just need to start concentrating on doing everything for themselves. Within my life I set a lot of goals from which I hope to achieve. One major goal is that of trying to drop my fifty meter freestyle time to twenty-five seconds. I work day in and day out on this and want to show myself I can do it. I don’t care what others think about my goals. I want to do this for myself and not for fame or money.

A person should not feel the need for fame or money because if they achieve a lifetime goal they should just be happy with themselves. Think about it I mean, how long does money last? Is fame really worth anything? I don’t think so. I mean, I know it would be nice to be recognized for doing something special or extraordinary but truthfully it won’t help you. Fame may make your day or make you smile for a couple of days but when you really think about it, it won’t make you happy for the rest of your life. Yet, if you do something for yourself and keep doing things for yourself you should always be happy with yourself.

Remember fame and money don’t last forever but one’s own happiness can last a lifetime and eternity. Therefore it is very obvious that the choice that should be made when trying to find motivation for a goal would be that of personal satisfaction rather than that of fame or money.

Why Essay #3 Received a Score of 4

This essay develops the point of view “that you should achieve a goal for your own satisfaction and not for that of fame.” The writer demonstrates competent critical thinking by presenting adequate evidence in a manner that demonstrates some progression of ideas from the short-term gain of “positive self-esteem,” supported by a scientific example from the writer’s life (“One major goal is that of trying to drop my fifty meter freestyle time...”), to the less specifically supported attainment of a “lifetime goal” with greater lasting value than money or fame. Facility in the use of language is evident (“Fame may make your day or make you smile for a couple of days but when you really think about it, it won’t make you happy for the rest of your life”) but inconsistent (“...it won’t make you happy for the rest of your life. Yet, if you do something for yourself and keep doing things for yourself you should always be happy with yourself”). The essay exhibits some variety in sentence structure (“A person should not feel the need for fame or money because if they achieve a lifetime goal they should just be happy with themselves. Think about it I mean, how long does money last? Is fame really worth anything?”). To merit a higher score, this essay needs to develop the point of view more insightfully and evenly, by providing further detailed evidence as support. This essay demonstrates adequate mastery and receives a score of 4.

Essay #4: This essay received a score of 3

Money and fame are nice but if you aren’t satisfied with your personal goals and achievement, then the money and fame don’t matter. Material things do not matter in the gist of life. People who are rich in one century are forgotten in the next. Mother Teresa accomplished more in a few years than most people do in a lifetime, even though she was not wealthy. If you are passionate for a cause, then the money and fame do not matter. Who remembers who the richest person in the world was in 1792? No one. However, I will always remember that in 1921, Susan B. Anthony fought for the right for a woman to vote but had enough passion about her cause that the rewards did not matter more than the accomplishment of her purpose.

Why Essay #4 Received a Score of 3

This response demonstrates developing mastery and some critical thinking by attempting to support a position that “Money and fame are nice but if you aren’t satisfied with your personal goals and achievement, then the money and fame don’t matter.” The writer limits focus to a central reason, “People who are rich in one century are forgotten in the next,” and supports this reason using two examples that are specific, though undeveloped and therefore inadequate (“Mother Teresa accomplished more in a few years than most people do in a lifetime...I will always remember that in 1921, Susan B. Anthony fought for the right for a woman to vote”). Although the writer demonstrates developing facility in the use of language (“Who remembers who the richest person in the world was in 1792? No one. However, I will always remember that in 1921, Susan B. Anthony fought for the right for a woman to vote”), vocabulary is sometimes weak and repetitive (“then the money and fame do not matter...then the money and fame do not matter...that the rewards did not matter”). To merit a higher score, this essay needs to exhibit stronger

critical thinking by providing further focused and detailed evidence to develop the point of view more effectively. Overall, this response remains in the inadequate category, earning a score of 3.

Essay #5: This essay received a score of 2

I have several ideas why it is personal satisfaction and not money or fame that urges people to succeed. My belief is after you succeed in something you do you feel good about yourself, you feel as you just successfully completed your mission that you made for yourself and it makes you feel good inside. For example, after an actor wins an Academy Award, I think he does not think about money or fame but instead he or she feels as she is great at her job and it makes that person complete and happy inside. Another example are authors who usually never got famous or wealthy when they are alive, but its rather the personal feeling of fulfillment of happiness that inspires them to write great stories. Those are several reasons why I think its personal feeling and satisfaction and not fame and fortune that inspires people to succeed.

Why Essay #5 Received a Score of 2

This response offers a seriously limited point of view (“its personal feeling and satisfaction and not fame and fortune that inspires people to succeed”) and demonstrates weak critical thinking by supporting this position with several brief examples (“after an actor wins an Academy Award...authors who usually never get famous or wealthy”) of situations in which “after you succeed in something you do you feel good about yourself.” However, this evidence consists of general statements that are insufficient to support the writer’s position (“I think he does not think about money or fame but instead he or she feels as she is great at her job and it makes that person complete and happy inside”). The response displays very little facility with language due to limited and repetitive vocabulary (“you feel good about yourself...you feel good inside...”). Consequently, to receive a higher score, this essay

needs to exhibit more skillful facility in the use of language as well as demonstrate stronger critical thinking by providing additional focused and specific evidence that will adequately develop the point of view. This response demonstrates little mastery and remains at the 2 score point.

Essay #6: This essay received a score of 1

My view of the idea that it is personal satisfaction rather than money or fame that motivates people to achieve is sometimes wrong because in sports some people do it for personal satisfaction because they love the game and some people do it for the money because it pays well. For example, in the NFL there are two types of people, one plays for the check and the other plays for the ring.

Why Essay #6 Received a Score of 1

Offering little evidence to support the writer’s point of view (“My view of the idea that it is personal satisfaction rather than money or fame that motivates people to achieve is sometimes wrong”), this response is fundamentally lacking. The sparse supporting evidence provided is weak (“in sports some people do it for personal satisfaction because they love the game and some people do it for the money”) and repetitive (“For example, in the NFL there are two types of people, one plays for the check and the other plays for the ring”). While the essay consists of only two sentences, one of these sentences displays flawed structure (“My view of the idea that it is personal satisfaction rather than money or fame that motivates people to achieve is sometimes wrong because in sports some people do it for personal satisfaction because they love the game and some people do it for the money because it pays well”). Thus, to merit a higher score, this essay needs to exhibit more control over sentence structure as well as demonstrate stronger critical thinking by providing further focused and detailed evidence to adequately develop the point of view. Overall, this essay demonstrates very little mastery in response to the writing task, and it is scored a 1.

Appendix B: A Sample Schedule for a Standard Setting Using Angoff with Mean Estimation to Set One Cut Score

Day 1

8:00–8:15 a.m.	Opening Remarks
8:15–9:00 a.m.	Overview of Goals and the Angoff with Mean Estimation Procedure
9:00–10:00 a.m.	Create Definition of the Just Minimally Competent Examinee
10:00–11:30 a.m.	Take Test and Review Responses
11:30 a.m. –12:15 p.m.	Review the Angoff with Mean Estimation Procedure and Practice Using the Method
12:15–12:30 p.m.	Complete Training Evaluation
12:30–1:30 p.m.	Lunch
1:30–2:00 p.m.	Answer Any Last Questions on the Angoff with Mean Estimation Procedure
2:00–3:30 p.m.	Make Round 1 Ratings
3:30 p.m.	Review Tomorrow’s Agenda and Dismiss (Remainder of afternoon will be needed to enter all ratings and compute results for the discussion)

Day 2

8:00–8:30 a.m.	Continental Breakfast
8:30–8:45 a.m.	Review of Goals and Agenda
8:45–10:00 a.m.	Provide Feedback on Round 1 Ratings and Facilitate Small-Group Discussions
10:00–11:30 a.m.	Make Round 2 Ratings
11:30 a.m.–12:30 p.m.	Lunch
12:30–1:30 p.m.	Provide Feedback on Round 2 Placements and Facilitate Small-Group Discussions
1:30–2:15 p.m.	Provide Impact Data and Facilitate Large-Group Discussion
2:15–3:00 p.m.	Make Round 3 Ratings and Complete Final Evaluation
3:00–3:30 p.m.	Final Debriefing and Dismissal

Appendix C: A Sample Schedule for a Standard Setting Using Angoff with Mean Estimation to Set Two Cut Scores

Day 1

8:00–8:15 a.m.	Opening Remarks
8:15–9:00 a.m.	Overview of Goals and the Angoff with Mean Estimation Procedure
9:00–10:00 a.m.	Create Definition of the Just Minimally Competent Examinee
10:00–11:30 a.m.	Take Test and Review Responses
11:30 a.m.–12:15 p.m.	Review the Angoff with Mean Estimation Procedure and Practice Using the Method
12:15–12:30 p.m.	Complete Training Evaluation
12:30–1:30 p.m.	Lunch
1:30–2:00 p.m.	Answer Any Last Questions on the Angoff with Mean Estimation Procedure
2:00–3:30 p.m.	Make Round 1 Ratings
3:30 p.m.	Review Tomorrow’s Agenda and Dismiss (Remainder of afternoon will be needed to enter all ratings and compute results for the discussion)

Day 2

8:00–8:30 a.m.	Continental Breakfast
8:30–8:45 a.m.	Review of Goals and Agenda
8:45–10:00 a.m.	Provide Feedback on Round 1 Ratings and Facilitate Small-Group Discussions
10:00–11:30 a.m.	Make Round 2 Ratings
11:30 a.m.–12:30 p.m.	Lunch
12:30–1:30 p.m.	Provide Feedback on Round 2 Placements and Facilitate Small-Group Discussions
1:30–2:15 p.m.	Provide Impact Data and Facilitate Large-Group Discussions
2:15–3:00 p.m.	Make Round 3 Ratings and Complete Final Evaluation
3:00–3:30 p.m.	Review the Need for a Second Cut Score and Review the Performance-Level Descriptors, Identifying How the Just

Minimally Competent Examinee for the Next Set of Ratings Differs from Those Just Completed (Assuming that all performance-level descriptors were created at one time on the first day)

3:30–5:00 p.m.

Make Round 1 Ratings

5:00 p.m.

Review Tomorrow’s Agenda and Dismiss (Remainder of afternoon will be needed to enter all ratings and compute results for the discussion)

Day 3

8:00–8:30 a.m.

Continental Breakfast

8:30–8:45 a.m.

Review of Goals and Agenda

8:45–10:00 a.m.

Provide Feedback on Round 1 Ratings and Facilitate Small-Group Discussions

10:00–11:30 a.m.

Make Round 2 Ratings

11:30 a.m.–12:30 p.m.

Lunch

12:30–1:30 p.m.

Provide Feedback on Round 2 Placements and Facilitate Small-Group Discussions

1:30–2:15 p.m.

Provide Impact Data and Facilitate Large-Group Discussion

2:15–3:00 p.m.

Make Round 3 Ratings and Complete Final Evaluation

3:00–3:30 p.m.

Final Debriefing and Dismissal

Appendix D: SAT Essay Prompt and Multiple-Choice Items for Writing— March 2005

From the 2004–2005 SAT Preparation Booklet



ESSAY

Unauthorized copying or reuse of
any part of this page is illegal.

ESSAY



ESSAY

Time — 25 minutes

Turn to page 2 of your answer sheet to write your ESSAY.

1. The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

Your essay must be written on the lines provided on your answer sheet—you will receive no other paper on which to write. You will have enough space if you write on every line, avoid wide margins, and keep your handwriting to a reasonable size. Remember that people who are not familiar with your handwriting will read what you write. Try to write or print so that what you are writing is legible to those readers.

You have twenty-five minutes to write an essay on the topic assigned below. **DO NOT WRITE ON ANOTHER TOPIC. AN OFF-TOPIC ESSAY WILL RECEIVE A SCORE OF ZERO.**

Think carefully about the issue presented in the following excerpt and the assignment below.

People who like to think of themselves as tough-minded and realistic tend to take it for granted that human nature is “selfish” and that life is a struggle in which only the fittest may survive. According to this view, the basic law by which people must live is the law of the jungle. The “fittest” are those people who can bring to the struggle superior force, superior cunning, and superior ruthlessness.

Adapted from S.I. Hayakawa, *Language in Thought and Action*

Assignment: Do people have to be highly competitive in order to succeed? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations

DO NOT WRITE YOUR ESSAY IN YOUR TEST BOOK. You will receive credit only for what you write on your answer sheet.

BEGIN WRITING YOUR ESSAY ON PAGE 2 OF THE ANSWER SHEET.

**If you finish before time is called, you may check your work on this section only.
Do not turn to any other section in the test.**

SECTION 5
Time — 25 minutes
35 Questions

Turn to Section 5 (page 5) of your answer sheet to answer the questions in this section.

Directions: For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

The following sentences test correctness and effectiveness of expression. Part of each sentence or the entire sentence is underlined; beneath each sentence are five ways of phrasing the underlined material. Choice A repeats the original phrasing; the other four choices are different. If you think the original phrasing produces a better sentence than any of the alternatives, select choice A; if not, select one of the other choices.

In making your selection, follow the requirements of standard written English; that is, pay attention to grammar, choice of words, sentence construction, and punctuation. Your selection should result in the most effective sentence—clear and precise, without awkwardness or ambiguity.

EXAMPLE:

Laura Ingalls Wilder published her first book and she was sixty-five years old then.

(A) and she was sixty-five years old then
 (B) when she was sixty-five
 (C) at age sixty-five years old
 (D) upon the reaching of sixty-five years
 (E) at the time when she was sixty-five

(A) (B) (C) (D) (E)

2. Inside famed actor Lily Langtry’s private railroad car were a drawing room with a piano, bath fixtures of silver, and there were draperies trimmed with Brussels lace.
- (A) there were draperies trimmed with Brussels lace
 (B) draperies trimmed with Brussels lace
 (C) trimmed with Brussels lace were draperies
 (D) the draperies were trimmed with Brussels lace
 (E) draperies trimmed with Brussels lace were there
3. Samuel Adams was by no means the first American to espouse the democratic cause, but he has been the first who conceived the party machinery that made it practical.
- (A) has been the first who conceived
 (B) had been the first who conceived
 (C) was the first having conceived
 (D) was the first to conceive
 (E) having been the first to conceive

4. The plans were made too hastily, without enough thought behind it.
- (A) too hastily, without enough thought behind it
 (B) too hasty, without enough thought behind it
 (C) too hastily, without enough thought behind them
 (D) too hasty, and there is not enough thought behind them
 (E) too hastily, and there is not enough thought behind it
5. Many psychologists do not use hypnosis in their practices, it is because they know very little about it and are wary of it as a result.
- (A) practices, it is because they know very little about it and are wary of it as a result
 (B) practices because they know very little about it and are therefore wary of it
 (C) practices for the reason that they know very little about it, with resulting wariness
 (D) practices because of knowing very little about it and therefore they are wary of it
 (E) practices, their knowledge of it being very little results in wariness of it
6. No two of the specimens was sufficiently alike to warrant them being called members of a single species.
- (A) was sufficiently alike to warrant them being called
 (B) was sufficiently alike to warrant the calling of them
 (C) was sufficiently alike to warrant their being called
 (D) were sufficiently alike to warrant the calling of them
 (E) were sufficiently alike to warrant calling them

GO ON TO THE NEXT PAGE 

7. My grandson thinks he can cook better than any other woman at the fair; and he has the blue ribbons to prove it.
- (A) My grandson thinks he can cook better than any other woman at the fair; and he
 - (B) My grandson thinks he can cook better than any woman at the fair, and he
 - (C) My grandson thinks he can cook better than any woman at the fair, consequently he
 - (D) To think he can cook better than any other woman at the fair, my grandson
 - (E) Thinking he can cook better than any other woman at the fair, my grandson
8. Differing only slightly from the Greeks were the Roman theaters, which were often freestanding rather than part of a hillside.
- (A) Differing only slightly from the Greeks were the Roman theaters, which
 - (B) Differing only slightly from Greek theaters, Roman theaters
 - (C) Differing only in the slightest from the Greeks were the Roman theaters, which
 - (D) The Greeks differed only slightly from the Romans, they
 - (E) The Greek theaters differed from the Roman theaters only slightly, where they
9. When chronological order is followed too mechanically, they are obscuring rather than clarifying important relationships.
- (A) When chronological order is followed too mechanically, they are obscuring rather than clarifying important relationships.
 - (B) When chronological order is followed too mechanically, it obscures rather than clarifying important relationships.
 - (C) Chronological order, if too mechanically followed, obscures rather than it clarifies important relationships.
 - (D) Chronological order, if followed too mechanically, obscures rather than clarifies important relationships.
 - (E) If you follow a too mechanical chronological order, it obscures rather than clarifies important relationships.
10. Small marine crustaceans known as krill are often fed to farm animals, but there is not much human consumption.
- (A) animals, but there is not much human consumption
 - (B) animals, but consumption is not done much by people
 - (C) animals but are rarely eaten by people
 - (D) animals, but eating them is rarely done by humans
 - (E) animals, but among people there is not much consumption
11. The educator's remarks stressed that well-funded literacy programs are needed if everyone is to gain the skills required for survival in society.
- (A) that well-funded literacy programs are needed if everyone is to gain
 - (B) that well-funded literacy programs needed in gaining
 - (C) there is a need of well-funded literacy programs for everyone will gain
 - (D) a need for well-funded literacy programs and everyone will gain
 - (E) why well-funded literacy programs being necessary for everyone in gaining
12. The Portuguese musical tradition known as *fado*, or "fate," has been called the Portuguese blues because of their songs that bemoan someone's misfortune, especially the loss of romantic love.
- (A) of their songs that bemoan someone's
 - (B) of their songs bemoaning their
 - (C) its songs bemoan
 - (D) the songs that bemoaned
 - (E) of how it bemoans their

GO ON TO THE NEXT PAGE 



The following sentences test your ability to recognize grammar and usage errors. Each sentence contains either a single error or no error at all. No sentence contains more than one error. The error, if there is one, is underlined and lettered. If the sentence contains an error, select the one underlined part that must be changed to make the sentence correct. If the sentence is correct, select choice E. In choosing answers, follow the requirements of standard written English.

EXAMPLE:

The other delegates and him immediately
A B C
accepted the resolution drafted by the
D
neutral states. No error
E

(A) ● (C) (D) (E)

13. Every year, toy manufacturers gather groups of
A
children into playrooms, observing their choices of
toys as predicting which new products will become
B C
the most popular. No error
D E

14. During the last fifty years, we come to take radio
A B C
communication for granted, but the mere suggestion
that we could communicate in such a fashion
D
must once have seemed outlandish. No error
E

15. The uncompromising tone of a recent city hall
A
ordinance concerning the blocking of emergency
B
vehicles in traffic jams carry a stern warning to
C D
motorists. No error
E

16. Formed by volcanic eruptions over the last five
A
million years, the Hawaiian Islands containing an
B
incredibly wide variety of species—many found
C
nowhere else on Earth. No error
D E

17. Because the owl is usually nocturnal plus being
A B
virtually noiseless in flight, it is seldom seen by the
C D
casual observer. No error
E

18. An economical and efficient recycling center
A
is accessible to the public, responsive to community
B
needs, and comply with current federal regulations
C
governing waste disposal. No error
D E

19. Jean Toomer was not only the author of *Cane*, a
A
novel whose publication has been viewed as marking
B C
the beginning of the Harlem Renaissance, but also
a respected advisor among Quakers. No error
D E

20. Election returns came in from upstate New York
A
quite rapid, but the results from New York City
B
were known even faster. No error
C D E

GO ON TO THE NEXT PAGE 

21. As we rely more and more on the Internet, your need
 A B
 for effective security planning and design to safeguard
 C
 data has increased. No error
 D E
22. The book is essentially a detailed and
 A
very well documented record of what happened
 B C
 to each of the protestors. No error
 D E
23. Experts agree that permanently modifying eating and
 A
 exercise habits rather than merely dieting for brief
 B
 periods are the key to controlling weight. No error
 C D E
24. The ability to control the plots of our dreams is
 A B
 a skill, researchers have shown, that we can learn
 C
 if you want to change recurrent dreams. No error
 D E
25. In swimming as to soccer, Evangelina proved time
 A B
 after time to be an abler competitor than Juanita.
 C D
No error
 E
26. The common cold is one of our most indiscriminate
 A
 diseases; it makes no distinction between
 B C
you and me, millionaires and paupers, or athletes and
 D
 couch potatoes. No error
 E
27. Like his other cookbooks, in his new book
 A
 Chef Louis offers lengthy explanations of what
 B
he considers to be basic cooking principles. No error
 C D E
28. Paul Ecke, flower grower and hybridizer, became
 A
known as "Mr. Poinsettia" after developing new
 B
 varieties of the flower and by pioneering it
 C
as a living symbol of Christmas. No error
 D E
29. Long thought of as a quiet, stuffy place
 A B
 where people just borrowed books, libraries
 C
have been changing their images dramatically
 D
 over the last few years. No error
 E
30. To understand twentieth-century economic practices,
 A
we must be sufficiently familiar with Keynesian
 B C
 theories, whether one agrees with them or not.
 D
No error
 E

GO ON TO THE NEXT PAGE 

Directions: The following passage is an early draft of an essay. Some parts of the passage need to be rewritten.

Read the passage and select the best answers for the questions that follow. Some questions are about particular sentences or parts of sentences and ask you to improve sentence structure or word choice. Other questions ask you to consider organization and development. In choosing answers, follow the requirements of standard written English.

Questions 30–35 refer to the following passage.

(1) Not many children leave elementary school and they have not heard of Pocahontas’ heroic rescue of John Smith from her own people, the Powhatans. (2) Generations of Americans have learned the story of a courageous Indian princess who threw herself between the Virginia colonist and the clubs raised to end his life. (3) The captive himself reported the incident. (4) According to that report, Pocahontas held his head in her arms and laid her own upon his to save him from death.

(5) But can Smith’s account be trusted? (6) Probably it cannot, say several historians interested in dispelling myths about Pocahontas. (7) According to these experts, in his eagerness to find patrons for future expeditions, Smith changed the facts in order to enhance his image. (8) Portraying himself as the object of a royal princess’ devotion may have merely been a good public relations ploy. (9) Research into Powhatan culture suggests that what Smith described as an execution might have been merely a ritual display of strength. (10) Smith may have been a character in a drama in which even Pocahontas was playing a role.

(11) As ambassador from the Powhatans to the Jamestown settlers, Pocahontas headed off confrontations between mutually suspicious parties. (12) Later, after her marriage to colonist John Rolfe, Pocahontas traveled to England, where her diplomacy played a large part in gaining support for the Virginia Company.

31. What is the best way to deal with sentence 1 (reproduced below) ?
- Not many children leave elementary school and they have not heard of Pocahontas’ heroic rescue of John Smith from her own people, the Powhatans.*
- (A) Leave it as it is.
 - (B) Switch its position with that of sentence 2.
 - (C) Change “leave” to “have left”.
 - (D) Change “and they have not heard” to “without having heard”.
 - (E) Remove the comma and insert “known as the”.

32. In context, which of the following is the best way to revise the underlined wording in order to combine sentences 3 and 4 ?
- The captive himself reported the incident. According to that report, Pocahontas held his head in her arms and laid her own upon his to save him from death.*
- (A) The captive himself reported the incident, according to which
 - (B) Since then, the captive reported the incident, which said that
 - (C) Consequently, the captive himself reports that
 - (D) It seems that in the captive’s report of the incident he says that
 - (E) According to the captive’s own report of the incident,

33. Which of the following phrases is the best to insert at the beginning of sentence 10 to link it to sentence 9 ?
- (A) Far from being in mortal danger,
 - (B) If what he says is credible,
 - (C) What grade school history never told you is this:
 - (D) They were just performing a ritual, and
 - (E) But quite to the contrary,

GO ON TO THE NEXT PAGE 

34. Which of the following best describes the relationship between sentences 9 and 10 ?
- (A) Sentence 10 concludes that the theory mentioned in sentence 9 is wrong.
 - (B) Sentence 10 adds to information reported in sentence 9.
 - (C) Sentence 10 provides an example to illustrate an idea presented in sentence 9.
 - (D) Sentence 10 poses an argument that contradicts the point made in sentence 9.
 - (E) Sentence 10 introduces a new source that confirms the claims made in sentence 9.
35. Which of the following would be the best sentence to insert before sentence 11 to introduce the third paragraph?
- (A) It is crucial to consider the political successes as well as the shortcomings of Pocahontas.
 - (B) The Pocahontas of legend is the most interesting, but the historical Pocahontas is more believable.
 - (C) If legend has overemphasized the bravery of Pocahontas, it has underplayed her political talents.
 - (D) To really know Pocahontas, we must get beyond myth and legend to the real facts about her private life.
 - (E) Perhaps we will never really know the real Pocahontas.
36. What information is most logical to add immediately after sentence 12 ?
- (A) How Rolfe and Pocahontas happened to meet and marry
 - (B) Details about other versions of the legend concerning John Smith
 - (C) Reasons for the confrontations between the Powhatans and the Jamestown settlers
 - (D) An account of Rolfe's life and work in Virginia
 - (E) A brief summary of the other public events in Pocahontas' life

STOP

If you finish before time is called, you may check your work on this section only.
Do not turn to any other section in the test.

SECTION 10

Time — 10 minutes

14 Questions

Turn to Section 10 (page 7) of your answer sheet to answer the questions in this section.

Directions: For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

The following sentences test correctness and effectiveness of expression. Part of each sentence or the entire sentence is underlined; beneath each sentence are five ways of phrasing the underlined material. Choice A repeats the original phrasing; the other four choices are different. If you think the original phrasing produces a better sentence than any of the alternatives, select choice A; if not, select one of the other choices.

In making your selection, follow the requirements of standard written English; that is, pay attention to grammar, choice of words, sentence construction, and punctuation. Your selection should result in the most effective sentence—clear and precise, without awkwardness or ambiguity.

EXAMPLE:

Laura Ingalls Wilder published her first book and she was sixty-five years old then.

- (A) and she was sixty-five years old then
- (B) when she was sixty-five
- (C) at age sixty-five years old
- (D) upon the reaching of sixty-five years
- (E) at the time when she was sixty-five

A B C D E

37. The sales assistant arranged the gems on the counter, he proceeded to tell us about the origins of each stone.

- (A) The sales assistant arranged the gems on the counter, he
- (B) The gems, which were arranged on the counter by the sales assistant, who
- (C) The gems were first arranged on the counter by the sales assistant, then
- (D) After arranging the gems on the counter, the sales assistant
- (E) The sales assistant, having arranged the gems on the counter, he

38. A whistle-blower is when an employee reports fraud or mismanagement in a company.

- (A) when an employee reports fraud or mismanagement
- (B) an employee who reports fraud or mismanagement
- (C) reporting by an employee of fraud or mismanagement
- (D) if an employee reports fraud or mismanagement
- (E) fraud or mismanagement being reported by an employee

39. After Eliza, the heroine of Shaw's *Pygmalion*, is transformed from a flower girl into a gentlewoman, she realizes that one's social class matters less than your character.

- (A) she realizes that one's social class matters less than your
- (B) she realizes that one's social class matters less than one's
- (C) then realizing that one's social class matters less than their
- (D) having realized how social class matters less than
- (E) there is her realization about how social class matters less than

40. Knowing the roots of words that are hard to spell helps students to become a better speller.

- (A) helps students to become a better speller
- (B) is helpful to students who want to be a better speller
- (C) helps students to become better spellers
- (D) is helpful to students in becoming a better speller
- (E) helps a student be better spellers

GO ON TO THE NEXT PAGE 



41. Most experts believe that young children's not being given physical affection, this interferes with their normal development.
- (A) young children's not being given physical affection, this interferes
 - (B) for young children who have had physical affection withheld from them, it interferes
 - (C) the failure at giving young children physical affection would interfere
 - (D) when withholding physical affection from young children, it interferes
 - (E) the withholding of physical affection from young children interferes
42. Electronic bulletin boards, combining the convenience of a telephone with the massive information storage capacity of a computer, present messages on diverse subjects as astronomy, artificial intelligence, and skydiving.
- (A) diverse subjects as
 - (B) diverse subjects that are
 - (C) subjects of such diversity as
 - (D) subjects as diverse as
 - (E) a subject as diverse as
43. Free from British rule after the American Revolution, a strong central government was an idea that many of the representatives attending the Constitutional Convention were wary of.
- (A) a strong central government was an idea that many of the representatives attending the Constitutional Convention were wary of
 - (B) the idea of a strong central government made wary many of the representatives attending the Constitutional Convention
 - (C) many of the representatives attending the Constitutional Convention were wary of a strong central government
 - (D) many representatives at the Constitutional Convention felt wary toward a strong central government
 - (E) many representatives at the Constitutional Convention, wary of a strong central government
44. Being cleaner and longer-burning compared with bituminous coal, anthracite was the first coal widely used in the United States for both domestic and industrial purposes.
- (A) Being cleaner and longer-burning compared with
 - (B) Both cleaner and more longer-burning compared to
 - (C) Cleaner and longer-burning than
 - (D) By burning longer and more clean than
 - (E) Cleaner as well as longer-burning, unlike
45. At graduation, the speaker assured us that our many courses in the liberal arts had prepared us equally well for the challenges of working and further study.
- (A) had prepared us equally well for the challenges of working and further study
 - (B) had prepared us equally well for the challenges of work and of further study
 - (C) has supplied the preparation for challenging work along with further study
 - (D) leaves us prepared for the challenges of work and further study both
 - (E) were the preparation for making the challenges of work or further study easier
46. Modern bluegrass songs, telling of love and despair and celebrating mountain beauty, reflect the genre's rural origins.
- (A) Modern bluegrass songs, telling of love and despair and celebrating mountain beauty,
 - (B) Modern bluegrass songs through their telling of love and despair and celebrating mountain beauty,
 - (C) Because modern bluegrass songs tell of love and despair and also celebrating mountain beauty, they
 - (D) With modern bluegrass songs that tell of love and despair and celebrate mountain beauty, they
 - (E) Telling of love and despair, modern bluegrass songs celebrating mountain beauty, and they also

GO ON TO THE NEXT PAGE 



47. The fruit fly is often used to study genetic mechanisms, because it reproduces rapidly scientists can observe the effects of experiments on several generations.
- (A) mechanisms, because it reproduces rapidly
 (B) mechanisms, since it reproduces rapidly,
 (C) mechanisms, since, with its rapid reproduction,
 (D) mechanisms; because it reproduces rapidly,
 (E) mechanisms; then rapid reproduction allows
48. Benin was the first sub-Saharan African country to experience a “civilian coup”: they were a regime that was dominated by the armed forces and obliged by citizens to implement democratic reforms.
- (A) they were a regime that was dominated by the armed forces and obliged by
 (B) they had been a regime that was dominated by the armed forces, when they were obliged to
 (C) it had a regime, armed forces dominating, but then were obliged to
 (D) armed forces dominated them until this regime were obliged by
 (E) a regime, dominated by the armed forces, was obliged by
49. This legend about Admiral Nelson, like other naval heroes, are based only partially on fact.
- (A) like other naval heroes, are
 (B) like those of other naval heroes, are
 (C) like other naval heroes, is
 (D) like legends about other naval heroes, are
 (E) like legends about other naval heroes, is
50. Bats and mosquitoes come out at twilight, and the bats would look for mosquitoes and the mosquitoes would look for people.
- (A) and the bats would look for mosquitoes and the mosquitoes would look
 (B) and the bats come to look for mosquitoes while the mosquitoes look
 (C) the bats look for mosquitoes and the mosquitoes are looking
 (D) the bats looking for mosquitoes while mosquitoes would look
 (E) the bats to look for mosquitoes and the mosquitoes to look

STOP

If you finish before time is called, you may check your work on this section only.
Do not turn to any other section in the test.

Appendix E: Scoring Guide for the SAT Essay

From the 2004–2005 SAT Preparation Booklet

<p>Score of 6</p> <p>An essay in this category is <i>outstanding</i>, demonstrating <i>clear</i> and <i>consistent mastery</i>, although it may have a few minor errors. A typical essay:</p>	<p>Score of 5</p> <p>An essay in this category is <i>effective</i>, demonstrating <i>reasonably consistent mastery</i>, although it will have occasional errors or lapses in quality. A typical essay:</p>	<p>Score of 4</p> <p>An essay in this category is <i>competent</i>, demonstrating <i>adequate mastery</i>, although it will have lapses in quality. A typical essay:</p>
<ul style="list-style-type: none"> effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position 	<ul style="list-style-type: none"> effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position 	<ul style="list-style-type: none"> develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position
<ul style="list-style-type: none"> is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas 	<ul style="list-style-type: none"> is well organized and focused, demonstrating coherence and progression of ideas 	<ul style="list-style-type: none"> is generally organized and focused, demonstrating some coherence and progression of ideas
<ul style="list-style-type: none"> exhibits skillful use of language, using a varied, accurate, and apt vocabulary 	<ul style="list-style-type: none"> exhibits facility in the use of language, using appropriate vocabulary 	<ul style="list-style-type: none"> exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary
<ul style="list-style-type: none"> demonstrates meaningful variety in sentence structure 	<ul style="list-style-type: none"> demonstrates variety in sentence structure 	<ul style="list-style-type: none"> demonstrates some variety in sentence structure
<ul style="list-style-type: none"> is free of most errors in grammar, usage, and mechanics 	<ul style="list-style-type: none"> is generally free of most errors in grammar, usage, and mechanics 	<ul style="list-style-type: none"> has some errors in grammar, usage, and mechanics
<p>Score of 3</p> <p>An essay in this category is <i>inadequate</i>, but demonstrates <i>developing mastery</i>, and is marked by ONE OR MORE of the following weaknesses:</p>	<p>Score of 2</p> <p>An essay in this category is <i>seriously limited</i>, demonstrating <i>little mastery</i>, and is flawed by ONE OR MORE of the following weaknesses:</p>	<p>Score of 1</p> <p>An essay in this category is <i>fundamentally lacking</i>, demonstrating <i>very little</i> or <i>no mastery</i>, and is severely flawed by ONE OR MORE of the following weaknesses:</p>
<ul style="list-style-type: none"> develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position 	<ul style="list-style-type: none"> develops a point of view on the issue that is vague or seriously limited, demonstrating weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position 	<ul style="list-style-type: none"> develops no viable point of view on the issue, or provides little or no evidence to support its position
<ul style="list-style-type: none"> is limited in its organization or focus, and may demonstrate some lapses in coherence or progression of ideas 	<ul style="list-style-type: none"> is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas 	<ul style="list-style-type: none"> is disorganized or unfocused, resulting in a disjointed or incoherent essay
<ul style="list-style-type: none"> displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice 	<ul style="list-style-type: none"> displays very little facility in the use of language, using very limited vocabulary or incorrect word choice 	<ul style="list-style-type: none"> displays fundamental errors in vocabulary
<ul style="list-style-type: none"> lacks variety or demonstrates problems in sentence structure 	<ul style="list-style-type: none"> demonstrates frequent problems in sentence structure 	<ul style="list-style-type: none"> demonstrates severe flaws in sentence structure
<ul style="list-style-type: none"> contains an accumulation of errors in grammar, usage, and mechanics 	<ul style="list-style-type: none"> contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured 	<ul style="list-style-type: none"> contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning

Essays not written on the essay assignment will receive a score of zero.

Appendix F: Answer Key for the SAT Reasoning Test Writing Section

From the 2004–2005 SAT Preparation Booklet

Writing

Section 1

Essay score

Section 5

Correct
Answer

2	B
3	D
4	C
5	B
6	E
7	B
8	B
9	D
10	C
11	A
12	C
13	B
14	B
15	C
16	B
17	B
18	C
19	E
20	B
21	B
22	E
23	C
24	D
25	B
26	E
27	A
28	C
29	B
30	B
31	D
32	E
33	A
34	B
35	C
36	E

Number correct

Number incorrect

Section 10

Correct
Answer

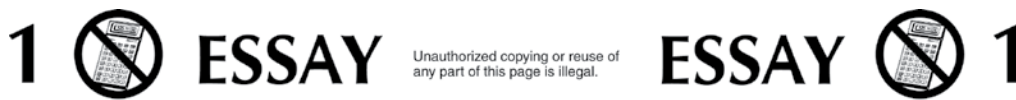
37	D
38	B
39	B
40	C
41	E
42	D
43	C
44	C
45	B
46	A
47	D
48	E
49	E
50	E

Number correct

Number incorrect

Appendix G: Training Essay Prompt and Multiple-Choice Items for Writing Section

From the 2005–2006 SAT Preparation Booklet



ESSAY
Time — 25 minutes

Turn to page 2 of your answer sheet to write your ESSAY.

1. The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

Your essay must be written on the lines provided on your answer sheet—you will receive no other paper on which to write. You will have enough space if you write on every line, avoid wide margins, and keep your handwriting to a reasonable size. Remember that people who are not familiar with your handwriting will read what you write. Try to write or print so that what you are writing is legible to those readers.

You have twenty-five minutes to write an essay on the topic assigned below. **DO NOT WRITE ON ANOTHER TOPIC. AN OFF-TOPIC ESSAY WILL RECEIVE A SCORE OF ZERO.**

Think carefully about the issue presented in the following excerpt and the assignment below.

Given the importance of human creativity, one would think it should have a high priority among our concerns. But if we look at the reality, we see a different picture. Basic scientific research is minimized in favor of immediate practical applications. The arts are increasingly seen as dispensable luxuries. Yet as competition heats up around the globe, exactly the opposite strategy is needed.

Adapted from Mihaly Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention*

Assignment: Is creativity needed more than ever in the world today? Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

DO NOT WRITE YOUR ESSAY IN YOUR TEST BOOK. You will receive credit only for what you write on your answer sheet.

BEGIN WRITING YOUR ESSAY ON PAGE 2 OF THE ANSWER SHEET.

**If you finish before time is called, you may check your work on this section only.
Do not turn to any other section in the test.**

SECTION 3
Time — 25 minutes
35 Questions

Turn to Section 3 (page 4) of your answer sheet to answer the questions in this section.

Directions: For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

The following sentences test correctness and effectiveness of expression. Part of each sentence or the entire sentence is underlined; beneath each sentence are five ways of phrasing the underlined material. Choice A repeats the original phrasing; the other four choices are different. If you think the original phrasing produces a better sentence than any of the alternatives, select choice A; if not, select one of the other choices.

In making your selection, follow the requirements of standard written English; that is, pay attention to grammar, choice of words, sentence construction, and punctuation. Your selection should result in the most effective sentence—clear and precise, without awkwardness or ambiguity.

EXAMPLE:

Laura Ingalls Wilder published her first book and she was sixty-five years old then.

(A) and she was sixty-five years old then
 (B) when she was sixty-five
 (C) at age sixty-five years old
 (D) upon the reaching of sixty-five years
 (E) at the time when she was sixty-five

A B C D E

3. Many ancient Eastern rulers favored drinking vessels made of celadon porcelain because of supposedly revealing the presence of poison by cracking.
- (A) because of supposedly revealing the presence of poison
 (B) for being supposed that it would reveal the presence of poison
 (C) because of being supposed to reveal poison in it
 (D) for it was supposed to reveal that there is poison
 (E) because it was supposed to reveal the presence of poison

2. The poet Claude McKay was a native of Jamaica who spent most of his life in the United States but writing some of his poems in the Jamaican dialect.
- (A) The poet Claude McKay was a native of Jamaica who spent most of his life in the United States but writing
 (B) Being that he was a Jamaican who spent most of his life in the United States, the poet Claude McKay writing
 (C) Although a native of Jamaica, the poet Claude McKay spent most of his life in the United States, he wrote
 (D) Although the poet Claude McKay spent most of his life in the United States, he was a native of Jamaica and wrote
 (E) Because he was a native of Jamaica who spent most of his life in the United States, the poet Claude McKay writing

GO ON TO THE NEXT PAGE 



The following sentences test your ability to recognize grammar and usage errors. Each sentence contains either a single error or no error at all. No sentence contains more than one error. The error, if there is one, is underlined and lettered. If the sentence contains an error, select the one underlined part that must be changed to make the sentence correct. If the sentence is correct, select choice E. In choosing answers, follow the requirements of standard written English.

EXAMPLE:

The other delegates and him immediately
 A B C
 accepted the resolution drafted by the
 D
 neutral states. No error
 E

A B C D E

4. The ambassador was entertained lavish by
 A
 Hartwright, whose company has a monetary
 B C
 interest in the industrial development of the
 D
 new country. No error
 E
5. Among the discoveries made possible by
 A B
 the invention of the telescope they found that
 C D
 dark spots existed on the Sun in varying numbers.
- No error
 E

GO ON TO THE NEXT PAGE 



Directions: The following passage is an early draft of an essay. Some parts of the passage need to be rewritten.

Read the passage and select the best answer for the question that follow. Some questions are about particular sentences or parts of sentences and ask you to improve sentence structure or word choice. Other questions ask you to consider organization and development. In choosing answers, follow the requirements of standard written English.

Question 6 is based on the following passage.

(1) My father has an exceptional talent. (2) The ability to understand people. (3) When I have a problem that I think no one else will understand, I take it to my father. (4) He listens intently, asks me some questions, and my feelings are seemingly known by him exactly. (5) Even my twin sister can talk to him more easily than to me. (6) Many people seem too busy to take the time to understand one another. (7) My father, by all accounts, sees taking time to listen as essential to any relationship, whether it involves family, friendship, or work. (8) At work, my father's friends and work associates benefit from this talent. (9) His job requires him to attend social events and sometimes I go along. (10) I have watched him at dinner; his eyes are fixed on whoever is speaking, and he nods his head at every remark. (11) My father emerges from such a conversation with what I believe is a true sense of the speaker's meaning. (12) In the same way, we choose our friends. (13) My father's ability to listen affects his whole life. (14) His ability allows him to form strong relationships with his coworkers and earns him lasting friendships. (15) It allows him to have open conversations with his children. (16) Furthermore, it has strengthened his relationship with my mother. (17) Certainly, his talent is one that I hope to develop as I mature.

6. Of the following, which is the best way to revise and combine sentences 1 and 2 (reproduced below) ?

My father has an exceptional talent. The ability to understand people.

- (A) My father has an exceptional talent and the ability to understand people.
 (B) My father has an exceptional talent that includes the ability to understand people.
 (C) My father has an exceptional talent: the ability to understand people.
 (D) My father has an exceptional talent, it is his ability to understand people.
 (E) Despite my father's exceptional talent, he still has the ability to understand people.

Appendix H: Angoff with Mean Estimation of Essay Ratings Training Record

Rater's Unique ID Number: _____ Group Number: _____

Please provide the average score to one decimal place that you believe the just minimally competent examinee would receive on the essay.

Training Round 1

1. Essay (between 1.0–6.0) _____

Clearly write the number of students out of a group of 100 just minimally competent examinees who you feel would answer each question correctly.

Training Round 1

2. _____

3. _____

4. _____

5. _____

6. _____

Appendix I: Training Evaluation Form

The purpose of this evaluation form is to obtain your feedback about the training you have received so far on the SAT Reasoning Test writing section and the Angoff with Mean Estimation standard-setting method. Your feedback will provide a basis for the facilitator to determine what information may need to be reviewed prior to the start of the actual standard-setting process.

Please complete the information below. Do not put your name or identification number on the form as we want your feedback to be anonymous.

Gender: Male Female

Race/Ethnicity:

Hispanic Asian African American White
 Other (Please specify: _____)

Years of Experience as a Faculty Member:

1–5 years 6–10 years 11–15 years
 16–20 years 20–25 years 25+ years

Affiliation:

2-year College 4-year College

Please read the following statements carefully. Place an **X** under *one* category (Strongly Agree, Agree, Disagree, or Strongly Disagree) to indicate the degree to which you agree with each statement.

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I understand the purpose of this workshop.				
2. The facilitator explained things clearly.				
3. I feel comfortable with the task of assigning ratings to each question.				
4. I feel comfortable with the task of assigning an average rating for the essay.				
5. I understand the concept of the just minimally competent examinee.				
6. I am ready to begin setting standards on the SAT Reasoning Test writing section.				

Appendix J: Angoff with Mean Estimation of Essay Ratings Record

Rater's Unique ID Number: _____

Group Number: _____

Please provide the average score to one decimal place that you believe the just minimally competent examinee would receive on the essay.

	Round 1	Round 2	Round 3
Essay (between 1.0–6.0)	_____	_____	_____

Clearly write the number of students out of a group of 100 just minimally competent examinees who you feel would answer each question correctly.

Round 1	Round 2	Round 3	Round 1	Round 2	Round 3
2. _____	_____	_____	27. _____	_____	_____
3. _____	_____	_____	28. _____	_____	_____
4. _____	_____	_____	29. _____	_____	_____
5. _____	_____	_____	30. _____	_____	_____
6. _____	_____	_____	31. _____	_____	_____
7. _____	_____	_____	32. _____	_____	_____
8. _____	_____	_____	33. _____	_____	_____
9. _____	_____	_____	34. _____	_____	_____
10. _____	_____	_____	35. _____	_____	_____
11. _____	_____	_____	36. _____	_____	_____
12. _____	_____	_____	37. _____	_____	_____
13. _____	_____	_____	38. _____	_____	_____
14. _____	_____	_____	39. _____	_____	_____
15. _____	_____	_____	40. _____	_____	_____
16. _____	_____	_____	41. _____	_____	_____
17. _____	_____	_____	42. _____	_____	_____
18. _____	_____	_____	43. _____	_____	_____
19. _____	_____	_____	44. _____	_____	_____
20. _____	_____	_____	45. _____	_____	_____
21. _____	_____	_____	46. _____	_____	_____
22. _____	_____	_____	47. _____	_____	_____
23. _____	_____	_____	48. _____	_____	_____
24. _____	_____	_____	49. _____	_____	_____
25. _____	_____	_____	50. _____	_____	_____
26. _____	_____	_____			

Appendix K: SAT Writing Composite Score Conversion Table

From the 2004–2005 SAT Preparation Booklet

Writing Multiple-Choice Raw Score	Essay Raw Score					
	1	2	3	4	5	6
-12	200	200	210	240	270	300
-11	200	200	210	240	270	300
-10	200	200	210	240	270	300
-9	200	200	210	240	270	300
-8	200	200	210	240	270	300
-7	200	200	210	240	270	300
-6	200	200	210	240	270	300
-5	200	200	210	240	270	300
-4	200	200	230	270	300	330
-3	210	230	250	290	320	350
-2	230	250	280	310	340	370
-1	240	260	290	320	360	380
0	260	280	300	340	370	400
1	270	290	320	350	380	410
2	280	300	330	360	390	420
3	290	310	340	370	400	430
4	300	320	350	380	410	440
5	310	330	360	390	420	450
6	320	340	360	400	430	460
7	330	340	370	410	440	470
8	330	350	380	410	450	470
9	340	360	390	420	450	480
10	350	370	390	430	460	490
11	360	370	400	440	470	500
12	360	380	410	440	470	500
13	370	390	420	450	480	510
14	380	390	420	460	490	520
15	380	400	430	460	500	530
16	390	410	440	470	500	530
17	400	420	440	480	510	540
18	410	420	450	490	520	550
19	410	430	460	490	530	560
20	420	440	470	500	530	560
21	430	450	480	510	540	570
22	440	460	480	520	550	580
23	450	470	490	530	560	590
24	460	470	500	540	570	600
25	460	480	510	540	580	610
26	470	490	520	550	590	610
27	480	500	530	560	590	620
28	490	510	540	570	600	630
29	500	520	550	580	610	640
30	510	530	560	590	620	650
31	520	540	560	600	630	660
32	530	550	570	610	640	670
33	540	550	580	620	650	680
34	550	560	590	630	660	690
35	560	570	600	640	670	700
36	560	580	610	650	680	710
37	570	590	620	660	690	720
38	580	600	630	670	700	730
39	600	610	640	680	710	740
40	610	620	650	690	720	750
41	620	640	660	700	730	760
42	630	650	680	710	740	770
43	640	660	690	720	750	780
44	660	670	700	740	770	800
45	670	690	720	750	780	800
46	690	700	730	770	800	800
47	700	720	750	780	800	800
48	720	730	760	800	800	800
49	720	730	760	800	800	800

This table is for use only with the writing section on the test in this booklet. Essays not written on the essay assignment will receive a score of zero.

Appendix L: Final Evaluation Form for the SAT Reasoning Test Writing Section

The purpose of this evaluation form is to obtain your feedback about the training you have received so far on the SAT Reasoning Test writing section and the Angoff with Mean Estimation standard-setting method. Your feedback will provide a basis for evaluating the training, methods, and materials in the standard-setting process.

Please complete the information below. Do not put your name or identification number on the form as we want your feedback to be anonymous.

Gender: Male Female

Race/Ethnicity:

Hispanic Asian African American White
 Other (Please specify: _____)

Years of Experience as a Faculty Member:

1–5 years 6–10 years 11–15 years
 16–20 years 20–25 years 25+ years

Affiliation:

2-year College 4-year College

Please read the following statements carefully. Place an **X** under *one* category (Strongly Agree, Agree, Disagree, or Strongly Disagree) to indicate the degree to which you agree with each statement.

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I understood the purpose of this workshop.				
2. The facilitator explained things clearly.				
3. I felt comfortable with the task of assigning ratings to each question.				
4. I felt comfortable with the task of assigning an average rating for the essay.				
5. I understood the concept of the just minimally competent examinee.				
6. I am comfortable with the recommended cut score.				
7. I would be comfortable defending this process to my peers.				

Appendix M: Dictionary of Terms

Angoff Method: One of many existing standard-setting methodologies. The Angoff method is an iterative process that focuses on individual questions and the probability that the just minimally competent examinee will answer each question correctly.

Authoritative Body: The group of individuals who will make policy decisions about the activities to take place in the standard-setting study, review the recommended cut score along with additional information, and have the final authority in deciding the final cut score to be adopted for use. The authoritative body does not participate in the standard-setting study.

Average Number Correct: The number of questions on average an examinee would be expected to answer correctly. Specifically, in this case, the number of questions on average the just minimally competent examinee would be expected to answer correctly. This number is determined by taking an average of the SME Number Correct values for that round of ratings.

Average Number Wrong: The number of questions on average that an examinee would be expected to answer incorrectly. This number is determined by subtracting the Average Number Correct from the total number of questions on the test.

Cut Score: The identified score at which anyone scoring at or above the score is considered to be a member of the higher group, while anyone scoring below the score is considered to be a member of the lower group.

Distractor: Any one of the possible answer choices from which an examinee may select that are presented with a multiple-choice question.

Distribution: The number of subject matter experts recommending each probability of a correct response from zero to one hundred percent.

Essay Raw Score: The average of all ratings for the just minimally competent examinee that were assigned by the subject matter experts in that round of ratings. This should be a number between 1.0 and 6.0.

Facilitator: A person with specific skills who is outside the process and does not have an immediate stake in the outcome of the standard setting. This person should be knowledgeable about the standard-setting process and possess the ability to manage the group of SMEs so that training is effective and the discussion is on topic and not dominated by any one person or group of people.

Formula Score Correction: A procedure used to remove the amount of error in a test score an examinee is expected to have achieved by guessing. In this example, each multiple-choice question has four possible answers (or distractors) so the probability of guessing the correct response is one-fourth or 25 percent. Therefore, the Average Number Wrong is divided by 4 to estimate the formula score correction.

Impact: The consequences or ramifications in terms of predicted numbers of examinees that will be assigned to each group if a specific cut score is used to separate examinees into groups or categories.

Just Minimally Competent Examinee: An examinee with sufficient knowledge and skills to score at the cut score and qualify for membership in the upper level, but just barely.

Mean Estimation: The process of providing an estimate for the average score a just minimally competent examinee would achieve on a test question scored along a scale rather than just correct or incorrect, such as the SAT Reasoning Test essay, which can be scored anywhere between 1.0 and 6.0.

Performance-Level Descriptors: A set of definitions that detail the knowledge or skill which members of each group or category should demonstrate. Each group or category should have distinct characteristics listed to enable the SMEs to easily distinguish one from the other. Performance-level descriptors should be written in positive terms, avoiding ambiguous words, and be measurable.

Scoring Guide: A set of guidelines used for scoring test questions that have a range of score points which the student may earn by demonstrating various levels of proficiency. The scoring guide will provide a description or list of characteristics for a response that would earn each score point. For the purposes of this document, the essay has a scoring guide (see Appendix E) that describes the characteristics indicative of a score of 1 through a score of 6.

SME Number Correct: The sum of the probabilities provided to each question by one SME. If 30 SMEs participate in the standard-setting study, then there will be 30 SME Number Correct values.

Standard Error of Judgment (SEJ): If a different panel of the same size and constituency were to be convened and trained in exactly the same manner, the resultant recommendation would most likely be slightly different. In fact, should many panels be convened there would be a distribution of recommendations with a mean and a standard deviation. The SEJ is an estimate of the standard deviation of a large number of panel recommendations. The SEJ is frequently used to adjust the one (and generally only one) panel recommendation. The authoritative body may want to drop (or raise) the recommendation by one or two SEJs. For example, dropping the panel's recommendation by one SEJ may be interpreted that it is unlikely that a panel of all possible judges would have set the standard below the final accepted standard.

Standard Error of Measurement (SEM): All tests are fallible and an examinee is not likely to exhibit exactly the same score if they take a different version of the test or test the next day, assuming no additional training/learning occur overnight. Should an examinee retest many times (assuming no learning or fatigue) there would be a distribution of scores for the examinee. The mean of this hypothetical distribution of scores for an examinee is called the examinee's true score—what the examinee would get if there was no error in the measurement process. The standard error of measurement is an estimate of the standard deviation of this hypothetical

distribution of scores. If the authoritative body believes the panel's recommendation represents the best standard they may still want to adjust the standard downward (or upward). An examinee whose true score is exactly at the recommended standard will fail 50 percent of the time because there is error in the measurement process. The authoritative body may decide that failing a qualified examinee is a worse error than passing an unqualified examinee. In this case, the authoritative body might lower the standard by one (or two) SEM such that examinees at or slightly above the panel's recommendation are not likely to fail due to errors of measurement. Obviously, the probability of examinees slightly below the panel's recommendation also increases, so lowering the risk of failing a qualified examinee increases the risk of passing an unqualified examinee.

Standard-Setting Process: A set of methodologies that may be used to establish a cut score to separate examinees into adjacent groups or categories. Typically the process consists of one or multiple standard-setting studies to recommend a cut score, and an authoritative body that meets to make critical decisions before the studies and decides the final cut scores by using the results of the studies and other information.

Standard-Setting Study: One part of a more comprehensive standard-setting process. The standard-setting study collects recommendations of the cut score placement from a panel of subject matter experts through an objective process supplemented with discussion and, at times, empirical data.

Subject Matter Expert (SME): Faculty members with subject matter expertise in the area for which a cut score will be set. SMEs are also expected to have knowledge of the performance and skills required for an examinee to succeed at the level in question. Every effort should be made to acquire a representative group of SMEs for the standard-setting study.

Writing Multiple-Choice Raw Score: The number of multiple-choice questions an examinee is expected to answer correctly without guessing. This number is estimated by subtracting the Formula Score Correction from the SME Number Correct.

