

A Guide to Conducting Standard Setting for the Next-Generation ACCUPLACER® Placement Tests Using a Bookmark Procedure

Table of Contents

INTRODUCTION	3
OVERVIEW OF THE PROCESS	4
CHARACTERISTICS OF STANDARD SETTING PANELS	5
MATERIALS	5
PRE-MEETING MATERIALS	5
MEETING MATERIALS.....	6
<i>Item Sets</i>	6
<i>Evaluation Forms</i>	7
<i>Data Collection Forms</i>	8
PROCEDURE	8
TRAINING.....	8
<i>Test-taking Experience</i>	8
<i>Performance Level Descriptors (PLDs)</i>	9
ROUNDS OF RATINGS	9
<i>Preparation</i>	9
<i>First Round</i>	11
<i>Second and Third Rounds</i>	11
CONVERTING ITEM NUMBERS TO SCALE SCORES	12
FINAL RESULTS.....	12
CALCULATING THE STANDARD ERROR OF JUDGMENT	13
DOCUMENTATION	13
CONCLUSION	13
REFERENCES	14
ATTACHMENT A: AGENDA	15
ATTACHMENT B: EVALUATION FORMS	17
ATTACHMENT C: RECORD FORMS	23

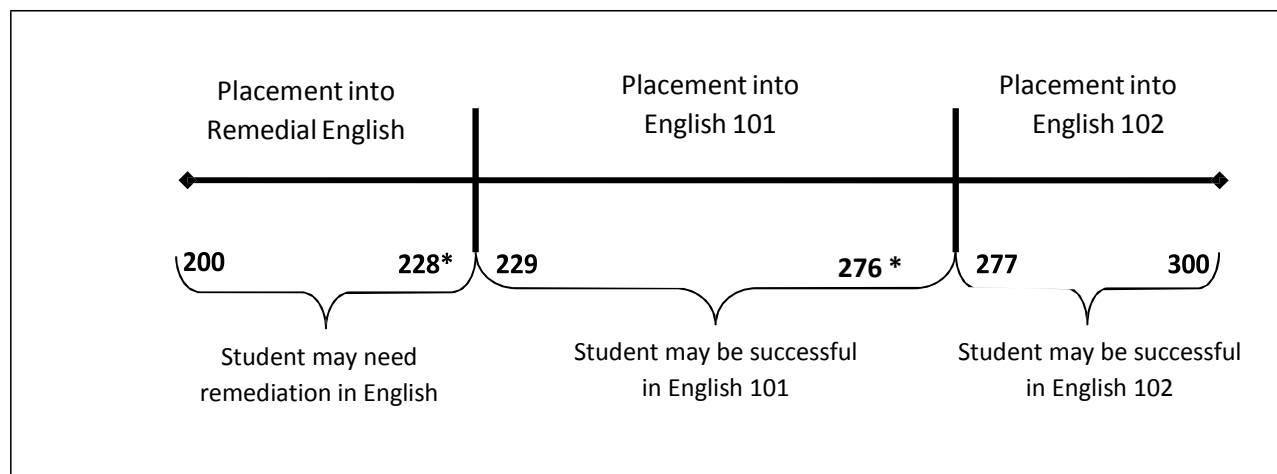
Introduction

ACCUPLACER® is a computerized-adaptive placement testing system delivered over the Internet that is used for assessing the knowledge and skills of incoming college students in a select number of content areas. The next-generation ACCUPLACER placement tests were designed to provide placement, advisement, and guidance information for students entering two- or four-year institutions of higher education. Each next-generation ACCUPLACER placement test has 20 items covering multiple content strands.

For use in course placement, institutions must set one or more cut scores on the next-generation ACCUPLACER placement score scale which ranges from 200—300. Each cut score represents the best estimate of the test score that separates the candidates who are minimally competent at the specified performance level from candidates who are not likely to be minimally competent. The performance level refers the categories into which examinees are to be placed by the standard setting process.

Figure 1 provides an example of categorizing examinees into three categories or, in the case of college placement, three courses. The cut scores are the points on the ACCUPLACER score scale that separate one course (or category) from the next.

Figure 1. Diagram of a Typical Use of Cut Scores for College Placement



*Cut score placements used in diagram are arbitrary and should not be interpreted as a recommendation for cut score placement.

This document provides information for one way of accomplishing this task, the Bookmark method (Mitzel, et al. 2001). This method is relatively easy to use and “perhaps the most *popular* method currently used to set performance *standards* on large-scale educational achievement test” (Cizek, 2012), and has withstood legal challenges (see, for example, Lewis, et al., 1999 & Mitzel, et al. 2001).

There are other standard setting methods which require different cognitive tasks of the standard setting panelists. There is no general consensus as to a preferred method and choosing a method may require effort or contacting a standard setting expert for advice. For other standard setting methods see Cizek (2001), and Cizek & Bunch (2007).

It is strongly recommended that a consultant or internal staff member with training and experience in standard setting be retained to facilitate and advise the institution on this process. The facilitator should have no stake in the process, be as unbiased as possible, and should not be allowed to vote or participate in discussions. The role of the facilitator is to provide necessary information to participants, train the participants, keep the discussion going but maintain reasonable timelines, maintain order, and ensure that every member of the panel has the opportunity to speak freely.

This paper is organized to describe the process from inception to documentation of the final decisions. The process which includes identifying appropriate panelists, standard setting materials, evaluation forms, analyses, and recommended documentation are described in some detail.

Overview of the Process

The Bookmark¹ method is typically a three-round standard setting process in which panelists work through a test booklet that has been re-ordered from easiest to hardest based on item difficulty. That is, test items from a prototypic version of the examination are ordered based on how well students have performed on the items. The items that students answered correctly most frequently (i.e., easier items) precede items that have been answered correctly less frequently (i.e., more difficult items). Panelists are asked to place a bookmark at the point in the ordered item booklet (OIB) at which they feel students have demonstrated sufficient knowledge and skills to be just minimally competent at each performance level in that subject area. When making decisions about bookmark placement, panelists are asked to keep in mind the spot where a barely qualified examinee has a 2/3 chance of getting the item correct, or where 2/3 of minimally competent examinees would answer a question correctly.

Panelists typically place the first set of bookmarks independently and then receive information on how the placements of their bookmarks compare with those of their peers. There is then a group discussion followed by a second round of bookmark placements. Finally, there is another group discussion followed by the third (last) round of bookmark placements. Feedback information is customarily provided to panelists between rounds of bookmark placements.

The process described above does not begin until the panelists have been trained, however. In standard setting, training starts with the panelists receiving information about the process and the assessments

¹ Any standard setting process uses a method that is either designed by the original authors of the method for that specific implementation, or an established or published method. Given the multitude of details in implementing a standard setting study, features of an established method are often modified to suit the specific needs and situations so long as the main aspects of the method is maintained and implemented with proper observances of best practices. Thus, what is described in this document is an implementation of a modified Bookmark standard setting method.

on which the standards will be set. This helps them come prepared and confident to participate in a process for which the most important thing that they bring are their background and their judgement. The rest of the training is provided during the standard setting meeting.

In a typical Bookmark standard setting process the meeting begins with an overview of the ACCUPLACER placement test on which cut scores are being set, the reasons that cut scores are necessary, the decisions that will be made based on the cut scores (e.g., assignment to specific classes) and an introduction to the Bookmark standard setting process. Panelists then take the ACCUPLACER test to gain familiarity with the types of questions asked and format of the test. Next, panelists create a set of performance level descriptors (PLDs) which are descriptions of the knowledge, skills, and abilities that would be exhibited by the just minimally competent examinee at each performance level. The process is then reviewed again and panelists fill out an evaluation form to indicate their level of comfort with regards to what is required of them and their comfort with the process to be followed. Once panelists have indicated that they are comfortable with the procedure the operational standard setting will begin. Three rounds of bookmark placement with group discussion and feedback information between the rounds are suggested as described above.

Characteristics of Standard Setting Panels

A standard setting panel of experts familiar with the subject and the student population should be convened. The size of the panel should be approximately 15 – 20 people, if possible. The more panelists the better, however groups of more than 20 can become unwieldy and should be avoided in most cases. Panelists should be selected based on their

- knowledge of the subject matter assessed,
- familiarity with students in the respective courses, and
- appreciation of the consequences of setting the cut points at each performance level.

In the interest of equity, representatives from diverse geographic regions and from gender and major racial/ethnic subgroups should be asked to participate. Attention should also be given to diversity of professional experience and, in multiple campus systems, representation from each campus.

Materials

Materials for standard setting include training materials and those that are integral part of the process. Some materials are also necessary to maintain security of the assessment given that test items used in standard setting are currently part of the item pools being administered to students.

Pre-Meeting Materials

Prior to the standard setting workshop, panel members should receive formal communication providing them a description of the purpose and procedure of the standard setting study. This communication may be in the form of a letter or an email and should include information on the next-generation ACCUPLACER placement test or which the cut scores are going to be set as well as information about the course for which the placement test is going to be used. Appropriate information may be provided as a

link embedded in the letter or email. The workshop agenda should also be sent to the panelists in advance and should be enclosed with the letter or attached to the email. A sample agenda is provided in this document as *Attachment A*.

Meeting Materials

At the standard setting workshop, panelists receive training materials, practice materials, and a set of operational materials, including operational test items. It is important to note that items must be kept secure by assigning panelists an individual identification number and giving them material marked with the same number. Each panelist is required to sign an affidavit of nondisclosure, check the material out and in each day, and be responsible for controlling all documents labeled with his or her ID number. Facilitators should monitor each room to ensure no materials leave the rooms, and all materials are all accounted for at the end of each day, and at the end of the standard setting process.

In addition to the live test items, panelists also receive forms to record their bookmark placement, a practice test (without answers marked), and an ordered item booklet containing items ordered in terms of difficulty. The materials are described more fully below and prepared for the panelists from original set of materials received from the College Board upon request.

Item Sets

A set of 50 items have been selected by the College Board for use in standard setting for each of the next-generation ACCUPLACER placement test. Upon request, these item sets are sent to institutions for the sole purpose standard setting using the Bookmarking procedure. For each placement test, the set of items is provided in two ways: one for training called the Training Set and one bookmark placement called the Ordered Item Booklet (OIB). A copy of each set needs to be made for each panelist, along with the confidentiality agreement that each panelist using the materials has to sign.

Requesting Test Items to Use in a Bookmark Standard Setting Procedure

Test items for use in standard setting are available from the College Board by completing the data request form in Appendix A of the document “Guidelines for the Release of Data” which can be found at the following location:

http://www.collegeboard.com/prod_downloads/research/RDGuideforReleaseData.pdf.

Your ACCUPLACER Outreach representative can answer any question you may have or direct you further in obtaining data. Please allow at least 6 weeks before you need the item sets to allow for time to complete all paperwork, obtain necessary signatures, and have the materials you request sent to you in time for review prior to your planned use date. All items used in setting standards for the **next-generation** ACCUPLACER placement tests are live and operational. Due diligence is always expected in the storage and use of any data provided by the College Board. The confidentiality agreement to be signed by each panelist should be kept on file and provided to the College Board upon request.

Training Set

Part of the training aims to make panelists familiar with the test. Taking the test will provide them the experience similar to those of students. The **next-generation** ACCUPLACER placement tests are

computer-adaptive and this presents a limitation in standard setting. Taking a test composed of items with the same representation as the computer-adaptive counterpart is a reasonable alternative. The Training Set is a test booklet where each page contains one item and the items are ordered by their accession number. Furthermore, the answer key is not printed on each page. A separate page listing the items in the same order providing the correct answer for each item is to be distributed to the panelists after they take the test.

Ordered Item Booklet (OIB)

The Ordered Item Booklet (OIB) contains the same items as the Training Set with each page contains one item. In the OIB, the items are arranged by difficulty where the least difficult item is presented first and the most difficult item is presented last. The ordering is based on the level of student ability that corresponds to the 2/3 probability of responding to the item correctly. This is referred to as the RP67 value of the item. The OIB contains all the information about the item that panelists need to complete the Bookmarking task. Each page of the ordered item booklet shows the item, along with any passage or graphic, and the possible responses with the correct response indicated. A conceptual representation of the OIB is presented graphically in *Figure 2*. A separate page listing the items in the same order provides the RP67 value of each item. This listing informs the panelists how difficult items are relative to each other.

Evaluation Forms

Panelists are required to complete evaluation forms throughout the process to gauge their understanding of the procedure and to collect various types of evidence for future reports (see *Attachment B* for copies of the evaluation forms). One purpose of an evaluation form is to collect data to demonstrate evidence of procedural validity of the standard setting process. As described in a recent textbook on standard setting “Information about the panelists’ level of satisfaction with the performance descriptors, training, standard-setting process, and final standards is an important piece of the evidence for establishing the validity of the performance standards.” (Hambleton, 2001, p. 104)

The facilitator should know as soon as possible if panelists are not satisfied with the level of training they received. Therefore, the first evaluation form should be given to the panelists at the end of the training portion of the session to gauge their current understanding of the process and their comfort level with the tasks they will be performing. The evaluation forms can be quickly analyzed giving the facilitator an opportunity to review with the panelists materials that may appear to be unclear. Once panelists indicate they are comfortable with the process and ready to proceed. It is time to begin the bookmarking process.

It is also a good idea to ask panelists to provide a summative evaluation addressing the procedural validity of the standard setting process. All forms should be submitted anonymously with only the demographics identified. Sample summary evaluation forms are provided in *Attachment B* also. The facilitator should feel free to modify these forms to meet their needs. Sometimes the summative form is extended to also ask about the general experience such as food, lodging, facilities, or other factors that may be of interest.

Data Collection Forms

After each round of standard setting the panelists identify the item that represents their bookmark with regards to each performance level. To do so the panelists identifies the item before and after the bookmark. A sample data collection form is provided in *Attachment C*.

Procedure

In the standard setting workshop panelists are instructed to set the required number of cut scores at the just minimally competent level for each performance level on the ACCUPLACER. That is, the panelists are asked to find the points in the ordered item booklet which define the knowledge and skills needed to be just minimally competent at each performance level of the ACCUPLACER. The following sections describe the training and the judgments they will make.

Training

On the first day, a general orientation session is held where the need for a criterion-referenced cut score is explained and why these panelists were chosen to provide recommendations for these passing scores². The facilitator introduces the Bookmark approach for setting criterion-referenced cut scores along with the expectations for the panel members' participation. Following the general orientation, time is provided for more in-depth and hands-on training on the Bookmark method, in addition to the creation of Performance Level Descriptors (PLDs).

Test-taking Experience

Prior to working with the ordered-item booklet, panelists should first have an opportunity to take the operational ACCUPLACER. The purpose of taking the test is to allow the panelists to experience the test under similar conditions to the student. If possible, have panelists take the computer adaptive test (CAT) version of ACCUPLACER. However, in the event that this is not possible, use the Training Set and ask panelists to answer the questions. Remember, no key should be provided at this point. As the panelists complete the test, the facilitator should pass out the answer key. The panelists then may discuss any item or items that they felt might be tricky. The panelists should be cautioned, however, that they will evaluate each item as it exists and not suggest changes. The items are operational items and have been seen by real candidates and the panelists should evaluate the items with regards to the standard setting method. This helps panelists start to familiarize with the test items and start to think about what it is important for students to know to be successful at different levels.

² A common set of questions posed by panelists is, "Why was I chosen?" and "How was the panel chosen?" It is advisable to address these questions directly to alleviate any anxiety the panelists may be feeling about their qualifications to perform the novel task of setting standards.

Performance Level Descriptors (PLDs)

A key component of standard setting is developing performance level descriptions (PLDs) that reflect what is intended as the standard for entry-level performance into each course. These PLDs define the benchmark that panelists use in the subsequent standard setting process by describing performance of those students who just barely meet the criteria to be categorized into each performance level needed for course placement using ACCUPLACER at that institution.

During standard setting, panelists are asked to think of students who fit the criteria of just minimally competent, since the cut scores are set at the lowest level or first evidence of acceptable performance. Time should be spent in the standard setting workshops discussing the just minimally competent students as defined by the PLDs. Research shows that when panelists reach a consensus regarding the characteristics of these borderline students, more defensible standards result (Impara, Giraud, & Plake, 2000; Mills, Melican, & Ahluwalia, 1991). Developing the PLDs also serves to calibrate panelists so that they have a similar picture of the just minimally competent student at each performance level in mind as they enter the bookmarking procedure.

An example of PLDs used for a College Level Examination Program (CLEP) exam is provided in *Table 1*. This example provides an important look at the level of specificity desired. This example does exhibit different levels of success in a single course and is used to determine whether credit for the course should be awarded. The underlying concept does generalize to the ACCUPLACER tests by substituting the grade level descriptions for the course prerequisite information to determine the minimal level of knowledge and skills needed for success in each course into which a student may be placed. In ACCUPLACER, the descriptors would focus on skills needed for minimally competency to be placed in each course rather than on grade in course as shown in *Table 1*.

Rounds of Ratings

Preparation

Before setting the bookmarks, the panelists perform a complete review of the Ordered Item Booklets (OIBs). During this review, they are instructed to think about two questions:

1. What does this item measure?
2. What makes it more difficult than the previous item?

Table 1. Example of Performance Level Descriptors for a Math Test.

<p>The Minimally Competent A-level Student</p> <ul style="list-style-type: none"> – 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 non-routine 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. <p>The Minimally Competent B-level Student</p> <ul style="list-style-type: none"> – 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. <p>The Minimally Competent C-level Student</p> <ul style="list-style-type: none"> – 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 multi-step 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 facilitator may want to lead this discussion or break panelists into small groups first so more people are engaged in the activity. The purpose of this exercise is for the panelists to gain a common understanding about what knowledge and skills each item requires. This stage is essential to setting a reasonable standard by ensuring that panelists have given consideration to every item no matter how difficult or facile it may be. In this phase of the workshop, however, panelists should be cautioned not to discuss the placement of the bookmarks but only to focus on individual items and the two questions posed above. If small groups are used in the discussion process, a table leader should be appointed for each group to lead discussion and record the main points of discussion for sharing with the larger group upon convening prior to the first round of ratings. Sharing with the larger group enables everyone to have heard the main points even if not in the smaller group where it was first discussed.

First Round

After they review the OIB, the panelists are asked to place their first set of bookmarks independently. They should be told to “place a bookmark after the hardest item for which two-thirds of those students with sufficient knowledge and skills to be just minimally qualified for each performance level on the ACCUPLACER would be able to answer correctly.” (See *Figure 2*) When it is desirable to set multiple cut scores on the same test title the panelists should begin with placing the bookmark focusing on one PLD at a time until all bookmarks have been placed into the ordered item booklet. They should also be reminded to consider the following pieces of information for each bookmark placement:

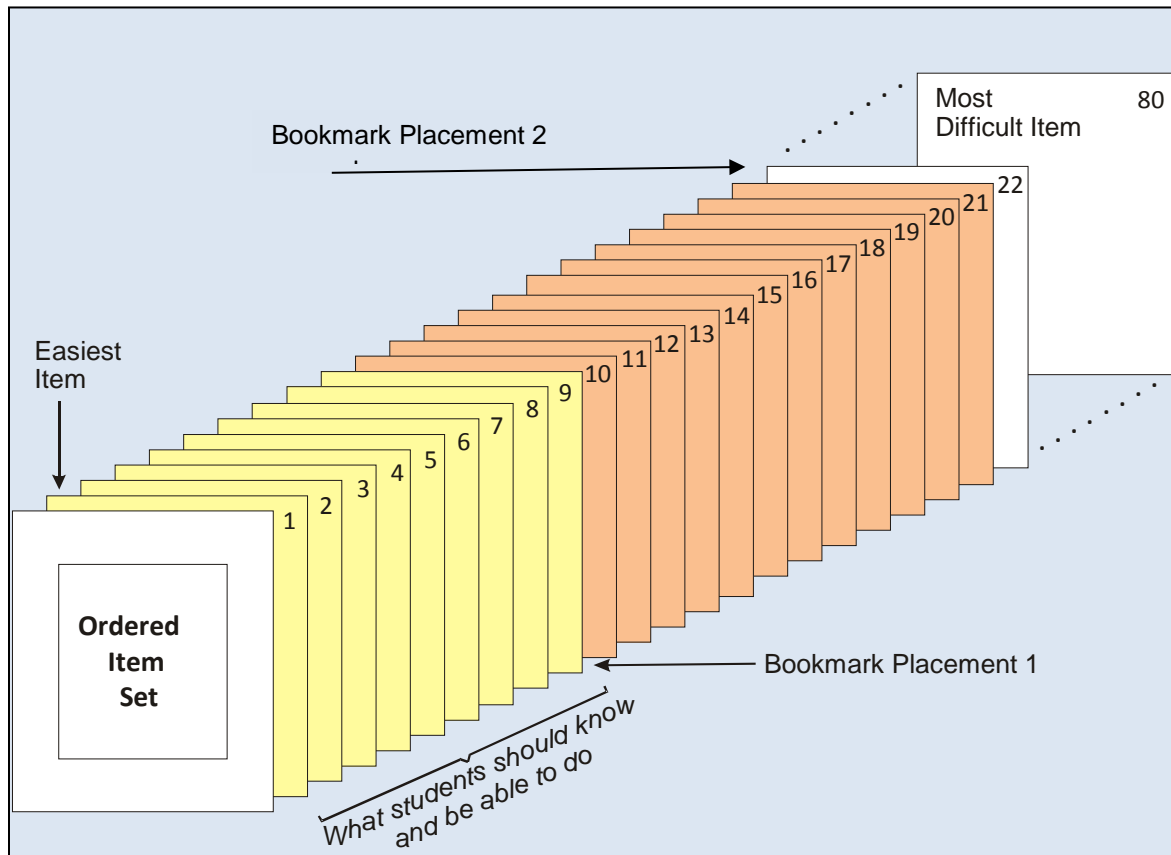
- Performance Level Descriptors
- Group discussions, which initially focused on the content and difficulty of each item and in later rounds focused on the amount of knowledge and skills required for students to be just minimally competent at each performance level intended for ACCUPLACER.

Second and Third Rounds

After completing each set of bookmark placements, the panelists recorded the item numbers on which they placed their bookmark. The facilitator should then enter the item number into a spreadsheet. The spreadsheet will calculate the median item number related to these item placements and highlighted the highest and lowest recommended item number. This analysis should be completed following each round of bookmark placements and panelists given the opportunity to discuss rationales for why the bookmark was placed at one item rather than at another favored by another panelist.

After the discussion, each panelist independently determines whether to move the benchmark(s) or leave it (them) where originally placed.

Figure 2. Ordered Item Booklet (OIB)



Converting Item Numbers to Scale Scores

Once the final item bookmark placements have been completed, find the median item number from the third round of ratings. [This](#) number should be sent to ACCUResearch@collegeboard.org with “Cut Score Request” on the subject line. You will then receive the corresponding ACCUPLACER scaled score based on the median item number you supplied.

Final Results

Typically, the results from a standard setting study are a recommendation for placement of a cut score and not considered final until an authoritative body has reviewed and signed off on the recommendation. This authoritative body may also choose, due to a variety of factors including a review of the summative evaluations by the panelists, to alter the recommended cut scores. Caution should be taken in altering the recommended cut scores with careful consideration to the rationale for why an adjustment is needed.

When policy makers choose to adapt the panel’s recommendation, they frequently use one of two statistics to aid in this decision. The first one is the standard error of judgment (SEJ) and the second is the standard error of measurement (SEM).

The SEJ is a measure of the degree of variability that might be observed if many similar panels of the same size were convened and went through the exact same standard setting exercise. The SEM describes the expected variation between the examinees' true score and their observed scores or may be conceptualized as the standard deviation of a candidate's scores should the candidate be administered the same test many times. The SEJ is idiosyncratic to the specific standard setting study and the equation is provided below. The SEMs can be located in the ACCUPLACER technical manual and available upon request from the College Board.

By bracketing the median passing score by 2 SEJs the 95 percent confidence interval for the average scores for panels can be determined. By bracketing the median passing score by 2 SEMs, the 95 percent confidence interval for candidate's scores can be determined. These standard errors can aid in identifying the appropriate passing score taking into consideration variance in the human judgments or the imprecision in the test itself.

Calculating the Standard Error of Judgment

The Standard Error of Judgment (SEJ) is produced in the following manner.

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 minus 1 (e.g., if the number of SMEs is 10 then divide by the square root of 9).

Documentation

Documenting the standard setting process and outcomes are important in the case that the institutions use of cut scores is ever challenged. All PLDs, record sheets, evaluation forms, etc. should be kept and the spreadsheet for each round of ratings used to calculate the median item number for each round should also be retained. Good documentation will allow you or your successor to answer the following questions:

- Who set the cut scores?
- How were the cut scores set? What was the process used?
- Was the process considered successful by the participants?
- Did the participants understand the process?
- Why was this score chosen instead of other possible scores?

Conclusion

Standard setting is an important part in any process that involves the categorization or placement of people into groups. The validity of the interpretation of the results of the testing experience requires that the test score be reliable and valid and that the cut scores be valid as well. This requires a standardized, defensible process for setting the standards. For this reason the College Board does

recommend that the process be managed by a professional trained in the methodology. This guide is intended to walk college administrators or test center personnel through the process of one of many existing methods for accomplishing this task. Many complexities exist that are not fully covered in this guide and it is strongly recommended that, before attempting to use this method, a trained facilitator or consultant should be retained to oversee the process.

It is a good practice to evaluate how well your cut scores are performing on a regular basis. It is suggested that institutions collect data at least every five years and use the Admitted Class Evaluation Service (ACES) to perform a validity study. ACES is a free service provided by the College Board to users of its products. Additional information on ACES can be found at <http://professionals.collegeboard.com/higher-ed/validity/aces>.

References

- Cizek, G. J. (Ed.) (2012a). *Setting performance standards: Foundations, methods, and innovations*. New York: Routledge.
- Cizek, G. J. (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cizek, G.J. & Bunch, M.B. (2007) *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: SAGE Publications.
- Hambleton, R. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*, (pp. 89-116). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Impara, J. C., Giraud, G., & Plake, B. S. (2000). The influence of providing target group descriptors when setting a passing score. Paper presented at the April 2000 meeting of the American Educational Research Association, New Orleans, LA.
- Lewis, D. M., Green, D. R., Mitzel, H. C., Baum, K., & Patz, R. J. (1999). *The bookmark standard setting procedure: Methodology and recent implications*. Manuscript under review.
- Mills, C. N., Melican, G. J., & Ahluwalia, N. T. (1991). Defining minimal competence. *Educational Measurement: Issues and Practice*, 10 (2), 7–10.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*, (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Attachment A: Agenda

Sample Agenda

Day 1

8:00 a.m. – 8:30 a.m.	Continental Breakfast
8:30 a.m. – 9:00 a.m.	Review Goals, Agenda, & the Bookmark Procedure
9:00 a.m. – 10:00 p.m.	Take Test and Review Responses
10:00 a.m. – 12:30 a.m.	Create Performance Level Descriptors
12:30 p.m. – 1:30 p.m.	Lunch
1:30 p.m. – 3:00 p.m.	Review Ordered Item Test Booklet
3:00 p.m. – 3:30 p.m.	Complete Evaluation Form / Answer Any Last Questions on the Bookmark Procedure
3:30 p.m. – 3:45 p.m.	Break
3:45 p.m. – 4:30 p.m.	Place Round 1 Bookmark
4:30 p.m. – 5:00 p.m.	Review Tomorrow’s Agenda and Dismiss

Day 2

8:00 a.m. – 8:30 a.m.	Continental Breakfast 8:30
a.m. – 8:45 a.m.	Review of Goals & Agenda
8:45 a.m. – 10:00 a.m.	Provide Feedback on Round 1 Placements and Facilitate Group Discussion
10:00 a.m. – 10:30 a.m.	Place Round 2 Bookmark
10:30 a.m. – 10:50 a.m.	Break
10:50 a.m. – 11:30 a.m.	Provide Feedback on Round 2 Placements and Facilitate Group Discussion
11:30 p.m. – 12:00 p.m.	Place Round 3 Bookmark and Complete Evaluation
12:00 p.m. – 12:30 p.m.	Final Debriefing, Evaluation, and Dismissal
12:30 p.m. – 1:30 p.m.	Lunch

Attachment B: Evaluation Forms

Evaluation of Training and Readiness for Beginning Cut Score Process

The purpose of this evaluation form is to obtain your feedback about the training you have received so far on ACCUPLACER and the Bookmark 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.				

Please add any comments you feel necessary to the back of this form.

Summative Evaluation Form

The purpose of this evaluation form is to obtain your feedback about the training you have received so far on ACCUPLACER and the Bookmark standard setting method. Your feedback will provide a basis for evaluation 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 cut score recommended.				
7. I would be comfortable defending this process to my peers.				

Please add any comments you feel necessary to the back of this form.

Alternative Summative Evaluation Form

Your feedback will provide a basis for evaluation 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 Teaching:

_____ 1-5 years _____ 6-10 years _____ 11-15 years
 _____ 16-20 years _____ 20-25 years _____ 25+ years

Affiliation:

_____ 2-year College _____ 4-year College
 _____ Other (Please specify: _____)

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. The ordered item booklet was easy to understand.				
4. The training on the Bookmark method was appropriate in giving me the information I needed to complete my assignment.				
5. The Performance Level Descriptors were useful in completing my assignment.				
6. I felt comfortable with the tasks.				

Bookmark Method for Next-Generation

7. I am comfortable with the cut scores recommended.				
8. I would be comfortable defending this process to my peers.				

Please rate the usefulness of the following materials or procedures in completing the standard setting process.

	Very Useful	Somewhat Useful	Not at all Useful
9. Taking the exam prior to the standard setting			
10. Creating the Performance Level Descriptors			
11. Reviewing the ordered item booklet at my table			
12. Group Discussions			

How influential was each of the following factors in placing your cut score?

	Very Influential	Somewhat Influential	Not Influential
13. The Performance Level Descriptors			
14. My perception of the difficulty of the items			
15. My personal experiences with students			
16. Discussions with other panelists			
17. The cut score placements of other panelists			

18. Were there any materials or procedures that became more (or less) influential in your placement of the cut score from one round to another? If so, which ones? Why?

How appropriate was the amount of time you were given to complete the different components of the standard setting practice?

	Too Much Time	About Right	Too Little Time
19. Training on the standard setting method			
20. Taking the test			
21. Scoring the test			
22. Discussing the items on the test			
23. Reviewing the ordered item booklet			
24. The discussion after Round 1			
25. The discussion after Round 2			

26. Please provide any additional comments you may have about this process.

35. How comfortable would you be defending this process to your peers?

____Very comfortable____Somewhat comfortable____Somewhat uncomfortable____Very uncomfortable

36. What could be changed to make you more comfortable defending this process?

37. Do you have any suggestions on how to improve the training and implementation of the standard setting workshop?

Thank You!

Attachment C: Record Forms

ACCUPLACER Standard Setting Record Sheet

Panelist ID# _____ Materials ID# _____ Test Name _____

For each round of standard setting record the item number for each cut you have set by writing the item number before and after your bookmark in the appropriate column below. For example, if you placed your bookmark between items 1 and 2 then you should write 1, 2.

Round	Cut score 1	Cut Score 2	Cut Score 3
1			
2			
3			

(You will need to modify this page by adding or deleting columns depending on the number of cut scores you plan to set. Rather than Cut Score 1, 2, 3 it may be helpful or less confusing to substitute the name or number of the course that students will be placed into with that cut score, i.e., English 98, English 99, English 101)