

# A Guide to Conducting Standard Setting for ACCUPLACER<sup>®</sup> Tests Using a Bookmark Procedure

---

## Table of Contents

<b>TABLE OF CONTENTS</b> .....	<b>2</b>
<b>INTRODUCTION</b> .....	<b>3</b>
<b>OVERVIEW OF THE PROCESS</b> .....	<b>6</b>
<b>CHARACTERISTICS OF STANDARD SETTING PANELS</b> .....	<b>7</b>
<b>MATERIALS</b> .....	<b>7</b>
PRE-MEETING MATERIALS .....	7
MEETING MATERIALS .....	7
<i>Test Question Sets</i> .....	8
Requesting Test Questions to Use in a Bookmark Standard Setting Procedure .....	8
Test Taking Booklet (TTB).....	8
Ordered Item Booklet (OIB) .....	9
<i>Evaluation Forms</i> .....	10
<i>Data Collection Forms</i> .....	10
<b>PROCEDURE</b> .....	<b>10</b>
TRAINING .....	10
<i>Test-taking Experience</i> .....	11
<i>Borderline Descriptors (BDs)</i> .....	11
ROUNDS OF RATINGS AND BETWEEN ROUND FEEDBACK .....	12
<i>Knowledge, Skills, and Ability (KSA) Review of the OIB</i> .....	12
<i>First Round</i> .....	14
<i>Between Rounds Feedback and Discussion</i> .....	14
<i>Second and Third Rounds</i> .....	15
CONVERTING QUESTION NUMBERS TO SCALE SCORES .....	15
FINAL RESULTS .....	15
CALCULATING THE STANDARD ERROR OF JUDGMENT .....	16
<b>DOCUMENTATION</b> .....	<b>16</b>
<b>CONCLUSION</b> .....	<b>16</b>
<b>REFERENCES</b> .....	<b>17</b>
<b>ATTACHMENT A: AGENDA</b> .....	<b>18</b>
<b>ATTACHMENT B: EVALUATION FORMS</b> .....	<b>20</b>
<b>ATTACHMENT C: RECORD FORMS</b> .....	<b>27</b>

## Introduction

ACCUPLACER® is a computerized-adaptive testing system delivered over the internet and is used for assessing the knowledge, skills, and abilities of incoming college students in a select number of content areas. The purpose of this document is to guide institutions in the conduct and implementation of an in-person standard setting using the Bookmark method for the following ACCUPLACER tests, three of which are English as a Second Language (ESL). The test titles are

- Reading
- Writing
- Arithmetic
- Quantitative Reasoning, Algebra, and Statistics (QAS)
- Advanced Algebra and Functions (AAF)
- ESL Reading Skills
- ESL Sentence Meaning
- ESL Language Use

ACCUPLACER tests were designed to provide placement, advisement, and guidance information for students entering two- or four-year institutions of higher education. For each ACCUPLACER test, students are required to respond to 20 questions (or items) covering multiple content strands, except for the Writing test for which students are required to respond to 25 questions.

For course placement using the Reading, Writing, and mathematics tests, institutions must set one or more cut scores on the ACCUPLACER score scale which ranges from 200 to 300. The score range for each of the ESL tests is 20 to 120. Each cut score represents the best estimate of the test score that separates the test takers who are minimally competent at the specified performance level from those who are not likely to be minimally competent. The performance levels (also called achievement levels) refer to the categories into which test takers are to be placed by the standard setting process. Figure 1 provides an example of categorizing test takers 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.

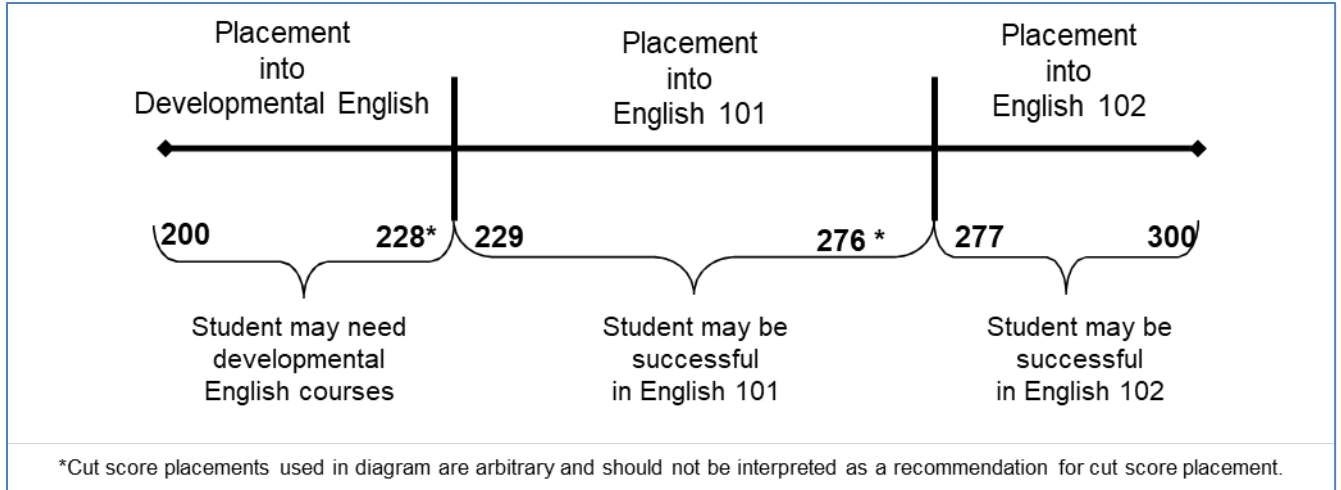


Figure 1: Diagram of a Typical Use of Cut Scores for College Placement

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 document 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, is described in some detail.

In the process described here, it is important to note that the materials will be provided electronically through a secure web application. Thus, it is necessary for the facilitator and each panelist to have a personal computer with internet connection to use during the standard setting process. A recommended physical setup for the standard setting implementation described here is presented in Figure 2.

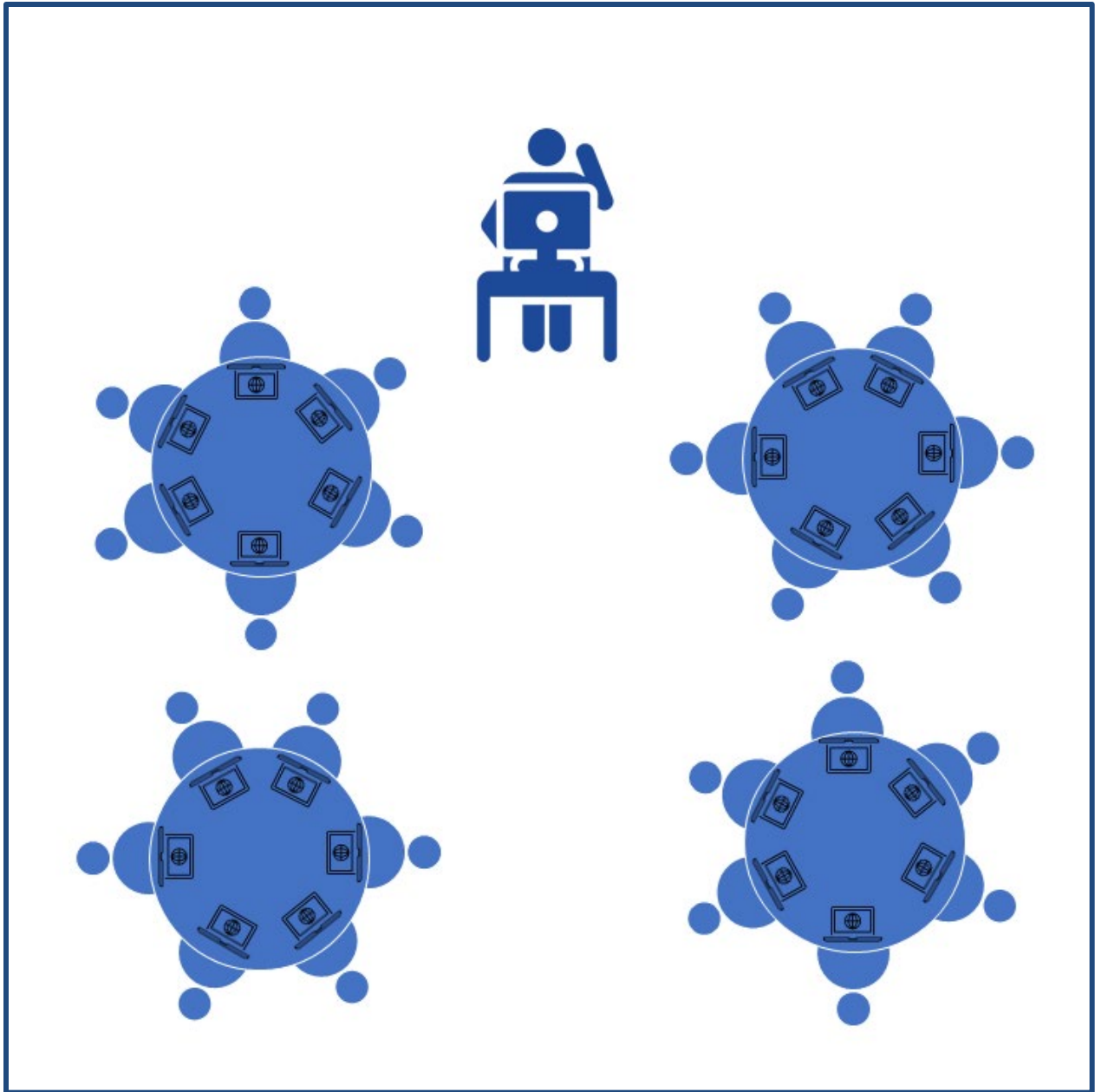


Figure 2: Physical Room Setup Recommended for the Standard Setting Process Described in this Guide

## Overview of the Process

The Bookmark<sup>1</sup> method is typically a three-round standard setting process in which panelists work through a test booklet with questions or items that have been re-ordered from the easiest to the hardest. That is, questions from a prototypic version of a test are ordered based on how well students have performed on the questions. The questions that students answered correctly most frequently (i.e., easier questions) precede questions that have been answered correctly less frequently (i.e., more difficult questions). 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, skills, and abilities 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 test taker has a 2/3 chance of responding correctly, or where 2/3 of minimally competent test takers 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 on which the standards will be set. This helps them come prepared and confident to participate in a process for which the most important things 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 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 for each performance level a Borderline Descriptor (BD) which is the description of the knowledge, skills, and abilities that would be exhibited by the just minimally competent test taker at that performance level. The process is then reviewed 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.

---

<sup>1</sup> 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.

## 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 20-24 people, if possible. The more panelists the better; however, groups of more than 30 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

Provision of proper materials is important to the success of any standard setting meeting. Some materials are received by the panelists in preparation for the standard setting. Materials used during the standard setting meeting include booklets containing live test questions as well as forms for collecting data. All materials necessary for implementing standard setting are described below. Secure materials will be provided through electronic access.

### Pre-Meeting Materials

Prior to the standard setting meeting, 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 ACCUPLACER test on 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 meeting agenda included in *Attachment A* of this document. The agenda may be modified to accommodate more flexible timing.

### Meeting Materials

At the standard setting meeting, panelists interact with ACCUPLACER test questions that are currently on the test. It is of paramount importance that test questions are kept secure. Materials that contain test questions will be provided in electronic form through WatchDox – a web application to provide secure read-only access to secure materials. WatchDox has been successfully used by College Board in test question review meetings for different testing programs as well as standard setting meetings (e.g., Bay & Duffy, 2000).

All standard setting panelists and meeting facilitator are required to sign an affidavit of nondisclosure prior to receiving access to secure materials. Once the affidavit is signed, College Board will create a secure WatchDox account for each standard setting participant. For training purposes, an advance preview of standard setting materials may be provided to standard setting facilitators upon request. In

the interest of keeping test materials secure, it is expected that the standard setting process is implemented behind closed doors with only the panelists and the facilitator having access to the live test questions. If for any reason there will be additional school personnel in the room during standard setting, those individuals will have to be identified to College Board and they will also have to sign the affidavit of nondisclosure.

In addition to booklets with live test questions, test descriptions will also be provided by College Board through WatchDox. Because test descriptions are not secure material, they will be accessible to the panelists and facilitator as soon as they successfully access their WatchDox accounts created for them by College Board.

During standard setting, panelists will also use forms to record their bookmark placements, as well as evaluation forms to help establish the procedural validity of the standard setting meeting. All the materials that will be used in standard setting are described more fully below. Only the materials containing live test questions and the test descriptions will be prepared by College Board. All other materials will be prepared by the institution with guidance and samples included in this document.

### **Test Question Sets**

A set of 50 questions have been selected by College Board for use in standard setting for each of the ACCUPLACER test. Upon request, these questions are made available electronically to institutions for the sole purpose of standard setting using the Bookmark procedure. For each test, the same set of questions is provided in two ways: one for training called the Test-Taking Booklet (TTB) and one for bookmark placement called the Ordered Item Booklet (OIB). A copy of each set is made available to each panelist during standard setting through their WatchDox account that will be created for them by College Board.

### ***Requesting Test Questions to Use in a Bookmark Standard Setting Procedure***

Test questions for use in standard setting are available from College Board by completing the electronic *Request for Data* form available at <https://collegeboard.tfaforms.net/69>. For additional details, see "*Guidelines for the Release of Data*":

[http://www.collegeboard.com/prod\\_downloads/research/RDGuideforReleaseData.pdf](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 question sets to allow for time to complete all paperwork, obtain necessary signatures, and have the materials you request prepared for your electronic access. All questions used in setting standards for ACCUPLACER tests are live and operational. The nondisclosure agreement (NDA) needs to be signed by each panelist and the facilitator and counter signed by College Board before electronic access to the materials is provided.

### ***Test-Taking Booklet (TTB)***

Part of the training aims to make panelists familiar with the test. Taking the test will provide them the experience similar to that of students. ACCUPLACER tests are computer adaptive and this presents a limitation in standard setting. Taking a test composed of questions with the same representation as the computer-adaptive counterpart is a reasonable alternative. The TTB is a test booklet where questions are ordered by their accession number. The answer key is not indicated on each question. A separate



page listing the questions in the same order providing the correct answer for each question is available. This page is referred to as the *Answer Key* and is to be opened only after completing the test. The intent is for the panelists to look at the correct answers only after they finished taking the test.

### Ordered Item Booklet (OIB)

The Ordered Item Booklet (OIB) contains the same questions as the TTB. In the OIB, the questions are arranged by difficulty: the least difficult question is presented first and the most difficult question is presented last. The ordering is based on the level of student ability that corresponds to the 2/3 probability of responding to the question correctly. This is referred to as the RP67 value of the question. The OIB contains all the information about the question that panelists need to complete the Bookmarking task. Each page of the OIB shows the questions, 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 3*. A separate page listing the questions in the same order provides the RP67 value of each question. This is referred to as the *Question List*. This listing informs the panelists how difficult questions are relative to each other.

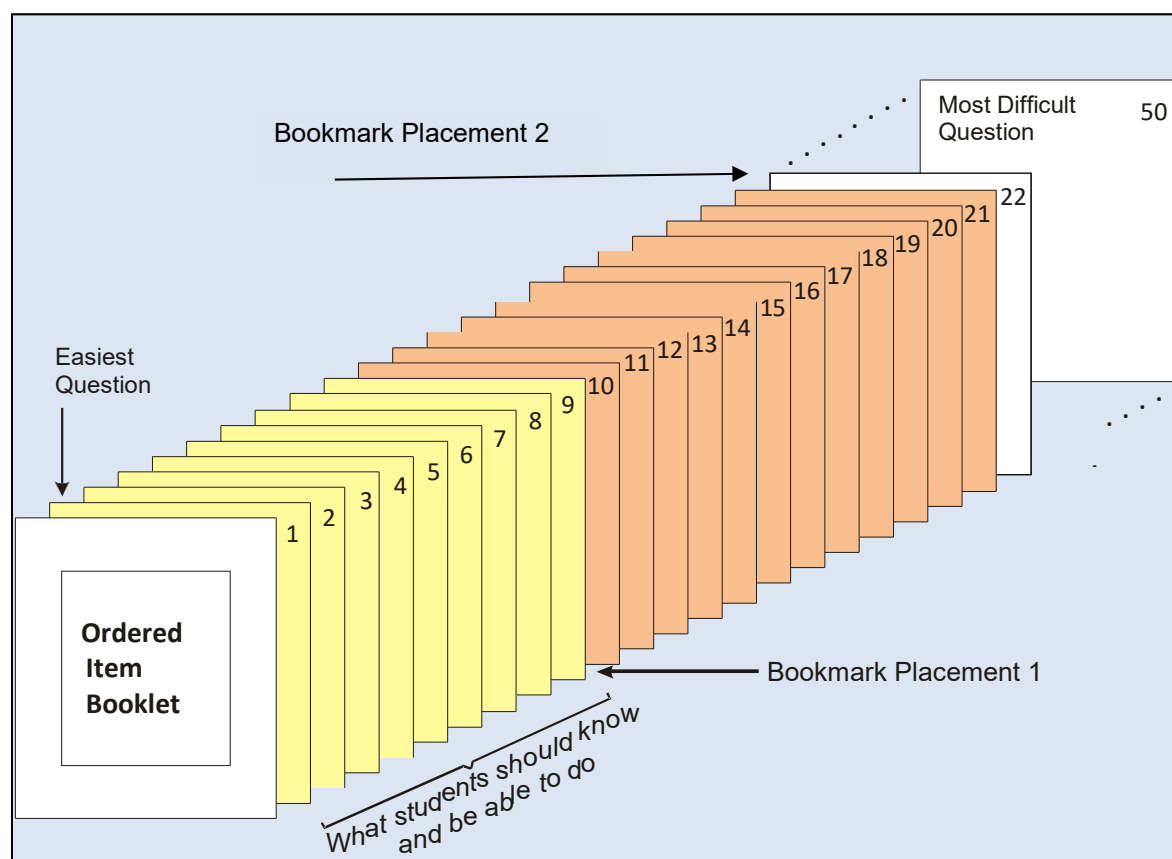


Figure 3: Ordered Item Booklet (OIB)

## 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 (*Training/Ready to Proceed 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 or instructions that may appear to be unclear. Once panelists indicate they are comfortable with the process and are ready to proceed, it is time to begin the bookmarking process.

It is also a considered best practice to ask panelists to provide an evaluation after every round of bookmarking and a *Summative Evaluation* addressing the procedural validity of the whole standard setting process. All forms should be submitted anonymously. Sample evaluation forms are provided in Attachment B. 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

At each round of standard setting, panelists place their bookmarks between questions that represent the delineation between adjacent performance levels. They are asked to record this placement by indicating the two questions between which they placed the bookmark on the OIB. To do so the panelists identify the questions before and after the bookmark. A sample data collection form is provided in *Attachment C*.

## Procedure

In the standard setting meeting panelists are instructed to set the required number of cut scores at the just minimally competent level for each performance level on an ACCUPLACER test. That is, the panelists are asked to find the points in the OIB which define the knowledge and skills needed to be just minimally competent at each performance level of the test. The following sections describe the training and 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 for the test for which panelists will be setting cut scores. The test description will be very useful in the discussion of the test. An explanation regarding the process for panelist selection and why

these panelists were chosen to provide recommendations for these passing scores<sup>2</sup> are customarily covered during the orientation. The facilitator then 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 familiarization of the test, in addition to the creation of Borderline Descriptors (BDs).

### **Test-taking Experience**

Prior to working with the OIB, panelists should first have an opportunity to take the operational ACCUPLACER test. The purpose is to give panelists insight into the test taking experience. If possible, have panelists take the computer-adaptive test (CAT) version of the ACCUPLACER test for which they will be setting standards. In addition to taking the CAT for the student experience, panelists will also have their first opportunity to become familiar with the test questions they will be interacting with by responding to the questions on the TTB. Remember, no key should be provided at this point. When panelists complete the test, they can look at the Answer Key which will be accessible in WatchDox. The panelists may then discuss any question or questions that they felt might be tricky. The panelists should be cautioned, however, that they will evaluate each question as it exists and not suggest changes. The questions are operational and have been seen by test takers and the panelists should evaluate the questions as they relate to the standard setting method. This helps panelists start to familiarize with the test questions and start to think about what is important for students to know to be successful at different levels.

### **Borderline Descriptors (BDs)**

A key component of standard setting is developing borderline descriptions (BDs) that reflect what is intended as the standard for entry level performance into each course. These BDs 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 meeting discussing the just minimally competent students as defined by the BDs. 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 BDs also serves to calibrate panelists so that they have a similar mental picture of the just minimally competent student at each performance level when they start the bookmarking procedure.

An example of BDs for a math test is provided in *Table 1*. This example provides an important look at the level of specificity desired. It exhibits different levels of success in a test and is used to classify student achievement into four different categories (Levels 1–4). The “threshold” descriptor for each level

---

<sup>2</sup> 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.

focuses on the knowledge, skills, and abilities minimally required for each level starting with level 2. The underlying concept does generalize to the ACCUPLACER tests by substituting the level threshold descriptions for the course prerequisite information to determine the minimal level of knowledge, skills, and abilities needed for placement in each course such that a student will have a reasonable probability of success in that course.

### **Rounds of Ratings and Between-Round Feedback**

The main part of a bookmark standard setting is placing bookmarks between test questions that signify delineation between adjacent performance levels. Bookmark placement is done iteratively three times and each iteration is referred to as a round of rating. Between rounds of rating, panelists receive information on the result of the previous round. Discussion of the results become basis for where they place their bookmarks in subsequent rounds. In preparation for the first round of ratings, panelists engage in a thorough review of the test questions in the OIB.

### **Knowledge, Skills, and Ability (KSA) Review of the OIB**

Before placing bookmarks, panelists perform a complete review of the OIB. Starting from the easiest question to the most difficult, panelists determine the knowledge skills, and abilities (KSAs) test takers need in order to answer each question correctly. They jot down the KSAs for each test question in the OIB in a “note to self” manner. For this review, it is helpful to think about what construct the particular question measures. It might also be helpful to discern what makes the question more difficult than the previous one.

The facilitator is encouraged to lead a discussion on a couple of selected questions (not necessarily the first two in the OIB) to model the thought process. After the discussion, panelists review the rest of the test questions in the OIB independently. Reviewing each question and determining the KSAs required to respond to it correctly makes for an efficient and effective first round of ratings.

Table 1. Example of Borderline Descriptors for a Math Test

<b>Achievement Levels Descriptors for Grade 11 Smarter Balanced Algebra*</b>
<b>Level 2 Threshold</b>
<ul style="list-style-type: none"> <li>• The student who just enters Level 2 should be able to:</li> <li>• Use linear equations in one and two variables and inequalities in one variable to model a familiar situation and to solve a familiar problem.</li> <li>• Explain solution steps for solving linear equations and solve a simple radical equation.</li> <li>• Use properties of exponents to expand a single variable (coefficient of 1) repeated up to two times with a nonnegative integer exponent into an equivalent form and vice versa, e.g., <math>x^2x^3 = xxx = x^2+3</math>.</li> <li>• Solve one-step linear equations and inequalities in one variable and understand the solution steps as a process of reasoning.</li> <li>• Represent linear equations and quadratic equations with integer coefficients in one and two variables graphically on a coordinate plane.</li> <li>• Recognize equivalent forms of linear expressions and write a quadratic expression with integer- leading coefficients in an equivalent form by factoring.</li> <li>• Add multi-variable polynomials made up of monomials of degree 2 or less.</li> <li>• Graph and estimate the solution of systems of linear equations.</li> </ul>
<b>Level 3 Threshold</b>
<p>The student who just enters Level 3 should be able to:</p> <ul style="list-style-type: none"> <li>• Create and use quadratic inequalities in two variables to model a situation and to solve a problem.</li> <li>• Write a quadratic expression in one variable with rational coefficients in an equivalent form by factoring, identify its zeros, and explain the solution steps as a process of reasoning.</li> <li>• Use properties of exponents to write equivalent forms of exponential functions with one or more variables with integer coefficients with nonnegative integer exponents involving operations of addition, subtraction, and multiplication without requiring distribution of an exponent across parentheses.</li> <li>• Solve a quadratic equation with integer roots in standard form.</li> <li>• Represent polynomial and exponential functions graphically and estimate the solution of systems of equations displayed graphically.</li> <li>• Understand that the plotted line, curve, or region represents the solution set to an equation or inequality.</li> <li>• Add and subtract multi-variable polynomials of any degree and understand that polynomials are closed under subtraction.</li> </ul>
<b>Level 4 Threshold</b>
<p>The student who just enters Level 4</p> <ul style="list-style-type: none"> <li>• should be able to:</li> <li>• Choose an appropriate equivalent form of an expression in order to reveal a property of interest when solving problems.</li> <li>• Solve a formula for any variable in the formula.</li> <li>• Provide an example that would lead to an extraneous solution when solving linear, quadratic, radical, and rational equations.</li> <li>• Use a variety of methods such as factoring, completing the square, quadratic formula, etc., to solve equations and to find minimum and maximum values of quadratic equations.</li> </ul>
<p>* Excerpted from Smarter Balanced Assessment Consortium’s Initial Achievement Level Descriptors and College Content-Readiness Policy (April 26, 2013) downloaded on December 12, 2019 from <a href="https://portal.smarterbalanced.org/library/en/mathematics-alds-and-college-content-readiness-policy.pdf">https://portal.smarterbalanced.org/library/en/mathematics-alds-and-college-content-readiness-policy.pdf</a></p>

## First Round

After the KSA review of the OIB, panelists review the next step, which is placing bookmarks on the OIB to set the cut scores. Because it is imperative that they have a good understanding of the standard setting process and its goals, they are asked to sign an evaluation form (*Training/Ready to Proceed Evaluation* in Attachment B) indicating that they are ready for the bookmarking task.

For placing their bookmarks, panelists are instructed to ask this question regarding each item starting from the easiest to the most difficult:

*If you had one student who perfectly matches the BD, would that student have a 67% chance of answering that question correctly?*

Alternatively, the question may be asked as follows:

*If you 100 clones of students perfectly matching the BD, would 67 of them be able to answer the question correctly?*

They should place the bookmark immediately preceding the question to which their answer is “no”. Note that the bases for their judgement in placing the bookmark are their understanding of the BD and the KSAs associated with the questions. Note that the placement of the bookmark divides the OIB questions into two groups: (1) questions that are easy enough for 2/3 of test takers who match the BD to answer correctly; and (2) questions too difficult for this expectation. (See *Figure 3*)

When multiple cut scores are to be set on the same test (i.e., the test is being used for placing students into different courses) the panelists should begin with placing the bookmark focusing on one BD at a time until all bookmarks have been placed into the OIB. Panelists’ bookmark placements are recorded, collected, and analyzed in preparation for the second round of ratings. A sample form for collecting data on bookmark placements is included in Attachment C. As panelists finish placing their bookmarks, they are asked to fill out the round 1 evaluation form. (See Attachment B.)

## Between-Round Feedback and Discussion

With the data collected, a spreadsheet is helpful in calculating the median bookmark placement. The median bookmark placement is where the panelists are considered to have placed their bookmark as a group. This information is shared with the panel as feedback from the first round of ratings. The lowest and highest bookmark placements are also shared with the panelists to indicate how far apart the bookmark placements were.

Prior to the next round of ratings, panelists engage in discussion that will enhance their judgments when they place their bookmarks. Panelists are encouraged to compare their own bookmark placement to the group placement and discern what could have caused any variance. The possible reasons are their understanding of certain questions with regard to the KSAs required to respond to them correctly. It is also possible that their understanding of the BD is different from that of the other panelists. In each case, panelists should have a discussion for the purpose of gaining a common understanding of the test questions and the BD. Panelists are encouraged to review their BD for the purpose of clarification.

Discussion in two stages is encouraged. After the feedback is provided, panelists discuss the results in

their table groups. After the table group discussion, the facilitator leads a whole group discussion where each of table group shares the important points made in their table group discussion. If edits are to be made to the BD, it should be discussed and agreed upon by the whole group.

### **Second and Third Rounds**

After the discussions, each panelist independently determines whether to move the bookmark(s) or leave it (them) where originally placed. Just like in the first round, the basis for bookmark placements in the second and third round is the panelists' understanding of the BD and the KSAs associated with the questions. Additionally, bookmark placements may also be based on the feedback and discussions just had.

Panelists fill out the Round 2 Evaluation Form as they finish their bookmark placement for round 2. Round 2 results are presented as feedback followed by table group and whole group discussion prior to the third and last round of ratings. Results from round 3 are the final results. Panelists fill out the Final Evaluation Form after they place their bookmarks in round 3.

### **Converting Question Numbers to Scaled Scores**

Once the final bookmark placements have been completed, find the median question number from the third round of ratings. This number should be sent to [ACCUResearch@collegeboard.org](mailto:ACCUResearch@collegeboard.org) with "Cut Score Request" on the subject line. You will then receive the corresponding ACCUPLACER scaled score based on the median question 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 test takers' true score and their observed scores and 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 are available upon request from 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 placement score taking into consideration variance in human judgments or imprecision in the test itself.

## Calculating the Standard Error of Judgment

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

1. Computing the standard deviation for the set of Round 3 judgments for the panelists
2. Dividing 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 is important in case the institution's use of cut scores is ever challenged. All BDs, record sheets, evaluation forms, etc. should be kept, and the spreadsheet for each round of ratings used to calculate the median bookmark placement 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, 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 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 College Board to users of its products. Additional information on ACES can be found at <https://aces.collegeboard.org/>.



## References

- Bay, L. & Duffy, L. (2020). *Texas Success Initiative Assessment 2.0 standard setting report: An online implementation*. The College Board: New York.
- 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, and Test Design
9:00 a.m. – 9:30 a.m.	Borderline Performance and the Bookmark Procedure
9:30 a.m. – 10:30 p.m.	Take the Test, Review your Responses, and Discuss Selected Test Questions
10:30 a.m. – 12:30 a.m.	Create Borderline Descriptors
12:30 p.m. – 1:00 p.m.	Lunch
1:00 p.m. – 3:00 p.m.	KSA Review of Test Questions
3:00 p.m. – 3:30 p.m.	Review the Bookmark Procedure and Complete the <i>Training/Ready to Proceed Evaluation Form</i>
3:30 p.m. – 3:45 p.m.	Break
3:45 p.m. – 4:30 p.m.	Place Round 1 Bookmarks
4:30 p.m. – 5:00 p.m.	Complete <i>Round 1 Evaluation Form</i> , Review Tomorrow's Agenda, and Dismissal

### Day 2

8:00 a.m. – 8:30 a.m.	Continental Breakfast
8:30 a.m. – 8:45 a.m.	Review of Day 2 Goals & Agenda
8:45 a.m. – 10:15 a.m.	Provide Feedback on Round 1 Bookmark Placements, Table Group Discussion, Whole Group Discussion, and Review Borderline Descriptors
10:15 a.m. – 11:00 a.m.	Place Round 2 Bookmarks and Complete <i>Round 2 Evaluation Form</i>
11:00 a.m. – 12:01 p.m.	Lunch
12:01 p.m. – 1:30 p.m.	Provide Feedback on Round 2 Bookmark Placements, Table Group Discussion, Whole Group Discussion, and Review Borderline Descriptors
1:30 p.m. – 2:00 p.m.	Place Round 3 Bookmarks and Complete <i>Summary Evaluation</i>
2:00 p.m. – 2:30 p.m.	Debriefing and Dismissal

## **Attachment B: Evaluation Forms**

## Training/Ready to Proceed Evaluation

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

Panelist ID#: \_\_\_\_\_

	Yes	No
I understand the purpose of the standard setting study.		
I have the technology facilities to help me do the standard setting tasks.		
I know how to go through test questions on WatchDox.		
I understand the steps I am to follow to make my standard setting judgments.		
I understand the concept of the borderline test taker.		
I am ready to complete my standard setting judgments.		

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

---



---



---

\_\_\_\_\_  
(Date)

\_\_\_\_\_  
(Signature)

\_\_\_\_\_  
(Print Name)

**Round 1 Evaluation**

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.

Panelist's ID#: \_\_\_\_\_

	<b>Strongly Agree</b>	<b>Agree</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
1. The facilitator explained things clearly.				
2. The use of the ordered item booklet (OIB) in the process was easy to understand.				
3. The training on the Bookmark method gave me the information I needed to complete my ratings.				
4. The Borderline Descriptors developed by the panel were accurate.				
5. The Borderline Descriptors were useful in helping me place my bookmark.				
6. I felt comfortable with the standard setting task.				
7. I understood how the questions were ordered in the OIB.				
8. It was easy to review questions on WatchDox.				

How influential was each of the following factors in placing your bookmark during Round 1?

	<b>Very Influential</b>	<b>Somewhat Influential</b>	<b>Not Influential</b>
9. Taking the test			
10. Identifying the KSAs of the questions in the ordered item booklet before making my ratings			
11. The Borderline Descriptors			
12. My perception of the difficulty of the questions			
13. My personal experiences with students			
14. The consequences of the test			

### Round 2 Evaluation

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.

Panelist's ID#: \_\_\_\_\_

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. The facilitator explained things clearly.				
2. The use of the ordered item booklet (OIB) in the process was easy to understand.				
3. The training on the Bookmark method gave me the information I needed to complete my ratings.				
4. The Borderline Descriptors developed by the panel were accurate.				
5. The Borderline Descriptors were useful in helping me place my bookmark.				
6. I felt comfortable with the standard setting task.				
7. I understood how the questions were ordered in the OIB.				
8. It was easy to review questions on WatchDox.				

How influential was each of the following factors in placing your bookmark during Round 1?

	Very Influential	Somewhat Influential	Not Influential
9. Taking the test			
10. Identifying the KSAs of the questions in the ordered item booklet before making my ratings			
11. The Borderline Descriptors			
12. My perception of the difficulty of the questions			
13. My personal experiences with students			
14. The consequences of the test			
15. My bookmark placement from Round 1			

## Final Evaluation Form

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

Panelist ID#: \_\_\_\_\_

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I understood the purpose of this meeting.				
2. The facilitator explained things clearly.				
3. Taking the test was helpful to me in understanding the student experience.				
4. I felt comfortable expressing my opinions.				
5. The use of the ordered item booklet in the process was easy to understand.				
6. The training on the Bookmark method gave me the information I needed to complete my ratings.				
7. The Borderline Descriptors were useful in making my judgements about bookmark placements.				
8. I felt we had sufficient discussion about questions, Borderline Descriptors and ratings.				
9. I felt comfortable with the Bookmark task.				
10. I am comfortable with the cut scores from the last round of bookmark placement.				
11. I would be comfortable defending this process to my peers.				

12. If you provided a rating of disagree or strongly disagree to question 10 or question 11 above, please explain why (use additional paper, if necessary).

---



---



---



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

	Very Useful	Somewhat Useful	Not at all Useful
13. Taking the test			
14. Creating the Borderline Descriptors			
15. Identifying the KSAs of the questions prior to placing the bookmark			
16. The Question List			
17. Small Group Discussions			
18. Large Group Discussions			

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

	Very Influential	Somewhat Influential	Not Influential
19. The Borderline Descriptors			
20. My perception of the difficulty of the questions			
21. My personal experiences with students			
22. Discussions with other panelists			
23. The Bookmark placements of other panelists			
24. The importance of the test to the test takers			
25. My experience taking the test			

26. Were there any materials or procedures that became more (or less) influential in your placement of the bookmark from one round to another? If so, which ones? Why? (Use additional paper if necessary.)

---



---



---



---

How appropriate was the amount of time you were given to complete the different components of the Bookmark process?

	<b>Too Much Time</b>	<b>Just Right</b>	<b>Not Enough Time</b>
27. Training on the Bookmark method			
28. Taking the test			
29. Identifying the KSAs for questions			
30. Placing the bookmarks			
31. Small Group Discussions			
32. Large Group Discussions			
33. Lunch			
34. Break			
35. Creating Borderline Descriptors			

36. What, if anything, 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 process?

---



---



---

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

---



---



---

## **Attachment C: Record Forms**

**ACCUPLACER Standard Setting Record Sheet**

Panelist ID#: \_\_\_\_\_

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

<b>Round</b>	<b>Cut score 1</b>	<b>Cut Score 2</b>	<b>Cut Score 3</b>
<b>1</b>			
<b>2</b>			
<b>3</b>			

(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)