

2003 Standard Setting for the WKCE

Wisconsin Knowledge and Concepts Examination (WKCE) was developed to assess student performance in core academic areas in accordance with the Wisconsin Model Academic Standards. For each grade-and-content area, the WKCE is designed to identify accomplishment at four different performance levels: Minimal Performance, Basic, Proficient, and Advanced. The staff of the Office of Educational Accountability (OEA) at the Department of Public Instruction (DPI) worked with its test vendor, CTB/McGraw-Hill, to establish proficiency-level cut scores for the Reading, Language Arts, Mathematics, Science, and Social Studies sections of the WKCE for grades 4, 8, and 10. The establishment of these performance levels was done through the use of the Bookmark Standard Setting Procedure (BSSP) in February 2003.

BSSP is a proprietary method of setting standards that was developed by CTB/McGraw-Hill for the specific purpose of simultaneously accommodating selected-response and constructed-response test formats, simplifying the cognitive complexity required of standard-setting judges and connecting test content with performance-level descriptors (Mitzel et al, 2001). The general framework for utilizing a bookmark standard-setting method provides participants with an ordered-item booklet. In this booklet, the items are ordered from the easiest to the most difficult, based on obtained difficulty scores from previous administrations of the items. The ordered-item booklet has one item per page, with the first item being the easiest, and the last item being the most-difficult. Each participant is instructed to place a bookmark at the location in the item booklet where the participant believes that a proficient student needs to master the knowledge, skills, and abilities contained in the items before the bookmark. Items preceding the participant's bookmark reflect content that all students at the given performance level are expected to know and to be able to perform successfully with a probability of at least 0.67. Conversely, these students are expected to perform successfully on the items behind the bookmark with a probability of less than 0.67. Three rounds of making bookmarks are typically conducted before results are summarized and before a final recommendation is made.

Participants

From February 9 through 12, 2003, a workshop was held in Madison, Wisconsin, with Wisconsin educators charged with the purpose of identifying the requirements for successful performance on the WKCE. Two hundred and forty panelists from 84 school districts attended the workshop; they included teachers from grades 3, 4, 7, 8, 9, and 10 as well as teachers of students with learning disabilities, of bilingual students and those with Title-I needs; administrators; curriculum experts; school-board members; guidance counselors; assessment coordinators; other curricular-area teachers, e.g., art and music; business leaders; parents, and other citizens from across the state.

Participants were nominated by professional organizations, by themselves, and by other referral sources. Each nominee was asked to provide qualifications and experiences. The

Department of Public Instruction selected the most-qualified panelists who also represented various areas of Wisconsin, who brought a rich variety of experience, and who collectively provided a balance of gender, ethnicity, and other factors.

Approximately 20 panelists participated in each grade-and-content area. Within each grade-and-content area, the participants were divided into three small groups. This was done to facilitate discussion between and among participants and to make the process more manageable for DPI and CTB/McGraw-Hill staff.

Method

The Bookmark Standard-Setting Procedure that was conducted for the WKCE consists of several steps that were designed to produce a recommendation for where the four proficiency level standards should be set. Each of these steps was essential in order to produce an informed decision that adheres to generally-recognized industry standards for validity and reliability.

After the initial orientation, the first step of the process was for participants to complete the operational form of the test. This was done to give participants a more-thorough understanding of the test content. Once this was done, participants were given copies of the examination items in order of difficulty (from easiest to most difficult) with one item per page. This Ordered-Item Booklet (OIB) was then analyzed, and participants were asked to identify what each item measured as well as why each item was more difficult than the preceding items. Following this, participants were provided training on how to place bookmarks. The bookmark placement required participants to move individually through the ordered-item booklet and to place a bookmark where the level of item difficulty for each category was surpassed. One bookmark was placed for each of the required cut points between Minimal Performance and Basic, between Basic and Proficient, and between Proficient and Advanced. Items preceding the mark reflected content that all students at the given performance level should have been expected to know and should have been able to do, defined as a likelihood of at least two-thirds.

At the beginning of the second stage of the BSSP, participants were provided with aggregate and disaggregate impact data based on the round-one placement of the bookmarks. Participants were then provided with the opportunity to discuss within their small groups the effect of placing the bookmarks at the selected locations and to identify the range of items between the lowest and highest bookmark for each proficiency level. Once this group activity was completed, participants were instructed to conduct the bookmark placement activity again individually.

The third stage of the BSSP was conducted in a fashion similar to that of the second stage. Participants were again provided with aggregate and disaggregate impact data based upon their recommendations and were given the opportunity to discuss their individual results within the small group. Participants were given the opportunity to place their individual bookmarks for a third and final time.

Final Recommendation

The final recommendations from the BSSP participants for each grade-and-content level were based on the final median-bookmark placements. These recommendations were then forwarded to the Wisconsin Technical Advisory Committee for their review and recommendations. This was then forwarded to the State Superintendent of Public Instruction, who adopted the recommendations of the standard-setting group and of the technical advisory committee.

References

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), IRT-based standard-setting procedures utilizing behavioral anchoring. Symposium conducted at the Council of Chief State School Officers National Conference on Large-scale Assessment, Phoenix, AZ.

Lin, Jie. (2004). The bookmark standard setting procedure: Strengths and weaknesses. The Centre for Research in Applied Measurement and Evaluation, University of Alberta: Edmonton, Alberta, Canada.

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: Erlbaum.

Bookmark Standard Setting Process:

States typically set cut scores on their assessments that characterize a student's performance on the assessment as at or above one of 3-4 performance levels (for example, basic, proficient and advanced). The method used to set these cut scores is called Standard Setting. There are several acceptable ways to set standards. Each depends on having participants in the standard setting who are very knowledgeable about the state's content standards and willing to help the state define the level of knowledge and skill expected of a student at each performance level articulated by the cut scores. An overview of standard setting and in particular CTB's Bookmark Standard Setting method is provided below.

Standard Setting

Standard Setting is a method of determining *cut scores* that correspond to performance levels. The standard setting process is usually implemented with a committee of educators since educators understand the state's content standards and the performance of students at each grade level and in each subject area. The *cut scores* that are determined during the Standard Setting procedure demarcate one performance level from another as is demonstrated in Figure 1. Figure 1 shows a state's performance levels separated by cut scores. During the Standard Setting, participants will set these cut scores by engaging in a structured conversation that includes discussion of content standards, performance levels, the test, and expectations for students.

Figure 1. State Performance Levels Separated by Cut Scores

For educators, this is a variation of a routine task. Teachers, for example, regularly separate students into performance levels, such as *A-Student*, *B-Student*, *C-Student*, *D-Student* and *F-Student*. The teacher assigns cut scores to each performance level; typical cut scores are 90% for *A-Student*, 80% for *B-Student*, 70% for *C-Student*, and 60% for *D-Student*. While these types of cut scores may be appropriate for classroom tests, their arbitrary and test-specific nature is not useful for a large-scale assessment. For a large-scale assessment, such as a state's subject area assessments, cut scores must take into consideration the difficulty of the test, as well as the knowledge, skills, and abilities expected of students to answer each test question.

Bookmark Standard Setting Procedure

To set valid, meaningful cut scores, a Standard Setting must be held. The Bookmark Standard Setting Procedure is a research-based method, recommended by CTB that can be used by committees of educators to establish cut scores on a state's assessments.

The Bookmark Standard Setting Procedure is a groundbreaking process developed by CTB/McGraw-Hill. Since its inception in 1996, over 28 states have implemented Bookmark to set cutscores on their large-scale assessments. The Bookmark Procedure typically includes training, 3 rounds of activities and discussion, and description writing.

Bookmark Materials: The Ordered Item Booklet (OIB) and its associated item map are key to the Bookmark Procedure. The OIB is constructed using items from the test. The items are ordered in terms of difficulty where the easiest item appears first and the hardest item appears last; this ordering is determined by student performance on the test. The Item Map gives detailed information about each item in the OIB and gives standard setting participants a place to record their thoughts about the items.

At the Bookmark Standard Setting Workshop

Over a three-day period, Bookmark standard setting participants engage in training, three rounds of structured discussions and ratings where they set cut scores on an assessment, and a writing session where they write descriptions of the recommended performance levels.

Round 1 of Discussion and Bookmark Placement: In small groups - typically tables of six to eight people - participants examine each item in the OIB, discussing what each item measures and what makes the item harder than those before it. After this discussion, each participant determines a cut score by placing a bookmark in the OIB according to his or her own judgement of what, for example, *Proficient* students should know and be able to do.

Rounds 2 and 3 of Discussion and Bookmark Placement: Participants then engage in two more rounds of bookmark placements. In Round 2, participants discuss the rationale behind their original bookmark placement with other participants at their table. In Round 3,

participants at all tables discuss their bookmark placements together. After each round of discussion, participants may adjust or maintain their bookmark placements. Impact data, that is the percentage of students in that state that would fall below each Bookmark, is introduced to participants during the third round. After the final round of bookmark placement, CTB Research calculates the group's recommended cut score by taking the median of all bookmark placements in the final round.

Description Writing: To complete the standard setting, participants write performance-level descriptors that reflect the final recommended cut scores. In small groups, participants examine the items before the bookmark and synthesize the content measured by those items. After two rounds of revisions, the performance-level descriptors represent a summary of the knowledge, skills, and abilities students must be able to demonstrate to enter each performance level.

Second Method Validation: For some states, a second method of standard-setting can be introduced at the end of the Bookmark standard-setting workshop in order to validate the recommended cut scores. CTB recommends using the contrasted groups method where teachers who have set the cut scores using the Bookmark methods then assign each of the students in their class rosters to a given performance level. CTB can check the actual performance of that student against the teacher's judgment. With cut scores being used for important decisions (e.g., WKCE is one factor in the promotion decision for students in Wisconsin), a second method gives the state more information upon which to set the final cut scores.

Final Decision on Cut Scores: The final determination of a state's cut scores is made by the individual or group with education policy-making authority within the state. The State Superintendent and/or the State Board of Education considers the recommendations of the standard setting committee and, oftentimes, review of those recommendations by a Technical Advisory Committee when establishing the final performance levels for a state's assessments. The more information the policymakers have when making this decision, the more defensible the final decision will be if the cut scores are challenged by a school, student or parent.

STATES THAT HAVE USED THE BOOKMARK STANDARD-SETTING APPROACH

STATE	ADDITIONAL INFORMATION
Alabama	
Alaska	
Arizona	
Arkansas	

California	
Colorado	
Connecticut	New CTB State - Bookmark being considered
Delaware	1999
District of Columbia	
Florida	
Illinois	2nd Grade Title I Test
Indiana	
Kentucky	Bookmark, Contrasted Groups, Jaeger-Mills
Louisiana	
Maryland	2003
Michigan	
Minnesota	Re-set HSGT standards
Mississippi	September 2002
Missouri	
Nevada	November 2002
New Mexico	
New York	Grades 4 & 8
North Dakota	
Oklahoma	1999
Pennsylvania	

South Carolina	
Tennessee	
Vermont	1998
Washington	
Wisconsin	
West Virginia	November 2002

Compiled by Dr. Karla Egan, Research Scientist, CTB/McGraw-Hill
September, 2002