



Setting Standards on the Prairie State Achievement Examination

ACT and the Illinois State Board of Education

Copyright © 2002, Illinois State Board of Education. All rights reserved.
The material in this booklet may be reproduced for educational and training purposes.
Any other use of this material must be approved by the Illinois State Board of Education.

Printed by the Authority of the State of Illinois
May 2002 2500 MY01741 No. 023

Setting Standards on the Prairie State Achievement Examination

Illinois students took the Prairie State Achievement Examination (PSAE) for the first time in spring 2001. On June 12–14, 2001, more than 100 educators and representatives of business from throughout the state met in Peoria to set standards for the five subjects assessed by the PSAE: reading, mathematics, writing, science, and social science. The goal of this process was to set cut scores to define the four levels of achievement—Exceeds Standards, Meets Standards, Below Standards, and Academic Warning—to be reported for the PSAE.

The standard-setting process was designed by ACT™ in collaboration with Illinois State Board of Education (ISBE) staff. A sample of 10,554 student PSAE records was available for producing feedback to panelists as part of the process. The overall standard-setting results were reasonable, and most panelists indicated that they were satisfied with the results and recommended their adoption.

In August 2001, State Board members approved the cut points that panelists recommended upon completion of the June standard setting. This report provides a brief overview of the process carried out at the June meeting and presents findings that support its validity.

Panelist Selection

Approximately 20 people served as panelists for each of the five PSAE subjects. Table 1 presents descriptive data for the panelists.

The majority of panelists (around 70 percent) were high school teachers in the subject for which they set standards. The category K–12 Educator includes a few individuals who were not teachers, such as school superintendents and curriculum directors. Representatives of the business community (17 percent) and higher education (12 percent) also served on each subject panel. Business panelists were employed in companies with ties to the subject—engineering for mathematics, pharmaceuticals for

Table 1: Descriptive Data for PSAE Standard-Setting Panelists

Subject	Number on Subject Panel	Panelist Type (number)			Gender (number)	
		K–12 Educator	Higher Education	Business Community	Males	Females
Reading	20	14	2	4	5	15
Mathematics	21	13	4	4	11	10
Writing	20	12	4	4	8	12
Science	19	16	1	2	9	10
Social Science	24	19	1	4	14	10
Total Number	104	74	12	18	47	57
Percent of Total		71	12	17	45	55

science, newspapers for writing, and so forth. Faculty who were teachers of introductory or freshmen courses or education courses in the subject represented higher education.

Panelists were selected to be as equal as possible with respect to gender: 45 percent males and 55 percent females. Regional representation was also a goal, and most subject panels included one or more panel members from each region of the state. Finally, teachers were drawn from schools of different sizes, and the distribution across the variable of school size was relatively even.

Facilitation of the Process

ACT staff, under the direction of a general process facilitator, served as process facilitators, while ISBE subject consultants served as subject facilitators. The general process facilitator directed preparation of outlines and presentation materials for each facilitator to ensure that each subject panel followed uniform procedures. Panelists received all instructions and training in general sessions to ensure uniformity.

The general process facilitator served as the process facilitator for reading, science, and social science panelists, and she trained additional process facilitators to work with panelists in mathematics and writing. The initial training lasted several days and extended over several weeks.

Mathematics was singled out for special facilitation because the mathematics portion of the PSAE contains a considerably greater number of assessment items than other subjects and thus required more time for item ratings. Writing was also singled out because the writing portion of the PSAE includes a writing sample, which required that writing panelists have slightly different training and use slightly different rating procedures. The entire process was presented to ISBE staff during an in-person meeting and reviewed in subsequent conference calls.

Subject and process facilitators met together for a full day of training shortly before the June 12–14 standard setting. The final day of training included review of all materials used in the process to make certain that all

facilitators understood the procedures to be implemented, the purposes of each, and the methods to use.

Training Panelists

Panelists began their training before they arrived in Peoria. They were sent a *Briefing Booklet* that described each session in the process, a preliminary agenda, and performance-level definitions and were asked to study the *Briefing Booklet* and performance-level definitions. Performance-level definitions are statements of what students *should* know and be able to do at each of the four performance levels: Exceeds Standards, Meets Standards, Below Standards, and Academic Warning.

On-site training began with a comprehensive overview to inform panelists about what they would be doing throughout the entire process. Presentations included varied formats, including printed text and slide shows.

As a reality check, panelists took the PSAE in their subject on the first day and scored their own test on the second day. Panelists were nearly unanimous in reporting that taking and scoring the test helped them understand what is expected of students.

On the second day, ISBE subject facilitators trained panelists in the Illinois Learning Standards and performance-level definitions. Panelists participated in exercises to help them understand the definitions more thoroughly, then developed descriptions of borderline performance for each performance level. This activity was aimed at helping panelists prepare for the item-rating process that required them to estimate how students at the lower borderline of each performance level *would* perform on each item.

Panelists requested more time than had been scheduled for this activity, and the agenda was adjusted to accommodate their needs. Panelists engaged in approximately 12 hours of training exercises and instructions to prepare them for the task of setting cut points. They spent additional time reviewing feedback data to prepare for the task of recommending final cut points.

Panelists' evaluations indicate that the training was successful for each subject group and panelist type. Panelists rated themselves on how clearly they understood the performance-level definitions and the concept of borderline performance. Their ratings were generally higher than 4.0 on a 5-point scale.

Rounds of Ratings

After panelists completed training in the performance-level definitions and reached agreement on descriptions of borderline performance for each level, they were ready for training in the rating procedures used to develop cut points. All panelists rated all PSAE items for their subject.

A modified Angoff method was used for the rating process: panelists estimated the percent of students at the lower borderline of each performance level who would correctly answer each test question. For the writing sample, panelists estimated the average score of students at the borderline of each performance level for each feature of writing. Panelists participated in the item-by-item rating process two times. They could change their ratings for any or all items at any or all performance levels.

Data analyses reveal that panelists made numerous changes in their item ratings, but the changes were not large. Reading panelists changed the largest proportion of their item ratings, and social science panelists changed the smallest proportion. Reading, writing, and science panel members changed approximately 60 percent of their ratings across all items for the three performance levels combined. Nevertheless, the average absolute change in estimates of student performance on the items was well below 10 percent, although reading panelists changed their ratings to a greater extent. Most of the changes by reading panelists were to decrease their estimates of student performance and lower the cut points from round 1 to round 2.

Panelists received feedback about their ratings after each round of ratings. Following the round 2 feedback, panelists were given a last opportunity to adjust their individual cut points. They were told that these would be averaged for all panel members in each group to produce the final cut points for each

subject. Panelists were allowed to take as much time as needed to complete their ratings, but all panelists completed the task within the amount of time estimated for this purpose.

When given the opportunity to make final adjustments in their own cut points for each performance level, most panelists made some changes, but the magnitude of absolute change was small. The differences in cut points from round 1 to round 2 were greater than between round 2 and the final recommendations. These findings are consistent with those from other standard-setting experiences in which approximately the same procedure was used. Panelists reported a high level of confidence in the ratings they had provided. They judged the process to be highly effective and the outcomes (cut points for each level and percent of students performing at or above each) to be both defensible and reasonable.

Feedback Data

Panelists were given several types of feedback data to inform them about their ratings and the consequences of their ratings. Cut points were reported on a pseudo-PSAE scale with a range of 500 to 670. This scale was used so that panelists would not compare cut points for the PSAE with those for other assessments, such as the Illinois Standards Achievement Test (ISAT) or the ACT Assessment[®]. Several types of feedback data were presented:

- Charts, graphs, and tables of data reporting information about the cut points,
- Percent of students who scored at or above each cut point for each subject,
- Location of each panelist's cut points for each level,
- Overall student performance on each test item in each subject, and
- Overall performance on a representative sample of about 30 test items of two students whose scores were right at the cut point set by round 1 ratings for each performance level.

Staff provided data to panelists after each round of ratings. Following the round 2 ratings, each panelist received data showing the percent of students who

scored at or above the cut points set by each panelist in the subject group. Those data aided the panelists in making recommendations for final cut points for each performance level in their subject.

Panelists were generally positive in their evaluations of the feedback data. They were impressed with the amount of information provided, and they found it helpful for evaluating their item ratings and cut points. They tended to rely on data about how students performed on each item and overall student performance on the PSAE more than on other feedback information. There were minimal differences in the understanding of and confidence in using feedback data by panelists of different types. The higher education panelists reported slightly lower levels of understanding and confidence than the other two types of panelists, but the differences were not great.

Evaluations

Panelists completed five process-evaluation questionnaires—one each day, with additional questionnaires regarding training in the feedback data on the second day and a final, overall process evaluation on the third day. Staff reviewed responses to questionnaires on site to monitor the process. Responses were analyzed for each subject and by panelist type and racial group for members of each subject panel. In general, the responses were quite positive. The average responses by panelists were higher than 3 on the 5-point evaluation scale, with a score of 5 being most positive.

Panelists’ responses to some questions were collected at several different stages in the process. These responses revealed that panelists’ understanding of the overall process and the key elements of the process (such as student performance relative to the performance-level definitions) increased and their confidence in performing the tasks increased from the first round of item ratings to the point when they recommended final cut points.

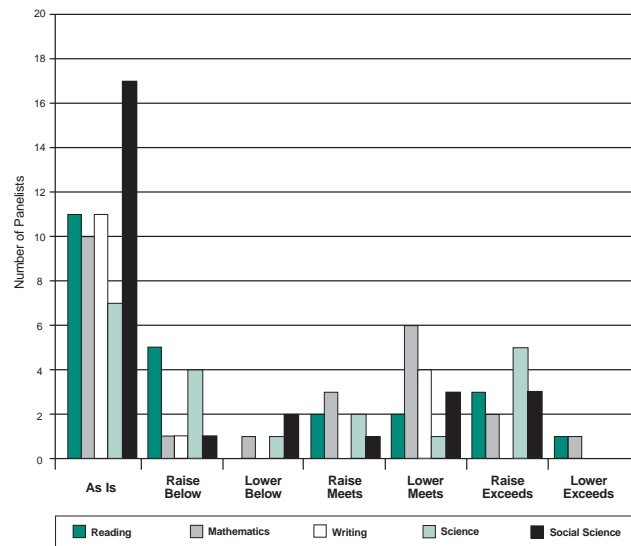
In the final process-evaluation questionnaire, panelists were asked if they would be willing to sign a statement recommending the cut points resulting from the process. Most panelists (93 percent) indicated they would *probably* or *definitely* sign such a statement.

Final Recommendations

Panelists’ final cut-point recommendations were averaged to compute the final cut points for each performance level in each subject. The consequences associated with the cut points were shared with panelists. Each panelist was then asked to complete a questionnaire evaluating those cut points and consequences. Panelists were told that their evaluations and recommendations would be reported to members of the State Board of Education to consider in reaching their final decisions to set standards for the PSAE.

The majority of panelists—nearly 60 percent—indicated that the outcomes were what they expected, and they recommended that the State Board adopt the final cut points as is. Other panelists had various responses, although panelists in each subject group tended to agree about the specific performance level(s) to change and the direction of change (raise or lower cut points). About 40 percent of the panelists recommended changes to the cut points. Figure 1 represents the recommendations by panelists in each subject group.

Figure 1: Recommendations by Subject Groups Regarding Final Cut Points



Data represented in Figures 2 and 3 show recommendations made by panelists of each of the three types. The graph in Figure 2 shows that a small number of panelists of each type recommended changes in the final cut points.

Figure 2: Recommendations to the State Board Regarding Final Cut Points, by Panelist Type

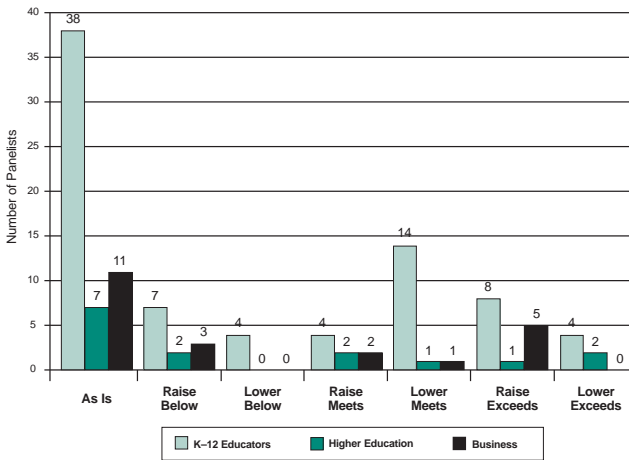
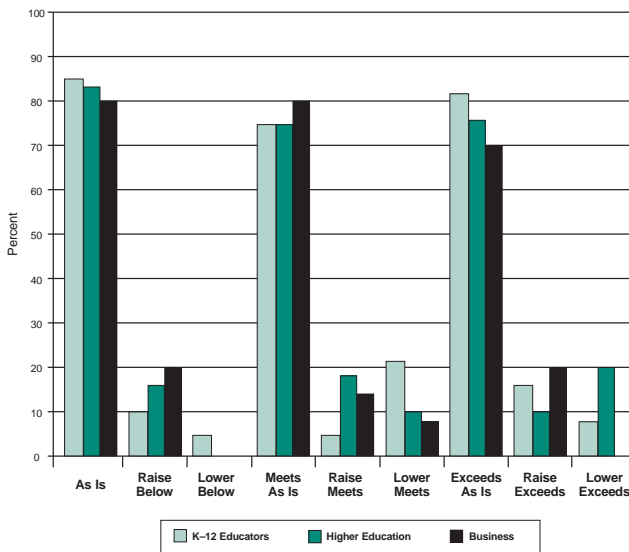


Figure 3: Percent of Panelists of Each Type Recommending Actions Regarding Cut Points for Each Performance Level



Teachers and other K–12 educators accounted for approximately 70 percent of the panel members, but it was helpful to look at the percent of panelists of each type making recommendations regarding the final cut points. The percent of panelists of each type recommending no changes in the cut point

for a performance level, raising the cut point, or lowering the cut point were computed and are graphed in Figure 3.

In general, the proportion of panelists of each type recommending the cut point for the Below Standards performance level as is was about even across the three types of panelists. For the Exceeds Standards level, the highest proportion of recommendations to keep the cut points as is was from K–12 educators and the lowest was from business panelists. Both figures show that recommendations to lower the cut points were more frequently made by K–12 educators and that recommendations to raise cut points were more frequently made by panelists who were not educators.

Review and Action by Members of the Illinois State Board of Education

Once the entire set of student performance data was scored and converted to the PSAE reporting scale, the percent of students scoring within each performance level was found to differ somewhat from those reported to panelists during the process. Table 2 includes both results:

- The results that were shared with panelists during the PSAE standard-setting process using cut points on the pseudo-PSAE scale (500–670) and percents of students in each performance level based on the small sample of student data, and
- The cut scores on the PSAE scale (120–200) adopted by State Board members and percents of students in each performance level based on all students who took the PSAE in spring 2001.

State Board members reviewed the recommendations made by the standard-setting panelists, the results based on the sample of 10,554 students, and the results based on the complete data set. After discussion, State Board members unanimously adopted the cut points that resulted from the standard-setting process in August 2001.

Table 2: Score ranges for performance levels and percents of student scores within each level based on (1) the pseudo-PSAE scale (500–670) and sample of 10,554 student records used in standard setting and (2) the final performance levels adopted by members of the Illinois State Board of Education and percents of student scores within each level on the PSAE scale (120–200).

Score Ranges on Pseudo-PSAE scale (500–670) and percent scores in each level (sample size = 10,554)								
Subject	Exceeds Standards		Meets Standards		Below Standards		Academic Warning	
	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)
Reading	639–670	13	604–638	49	574–603	33	500–573	5
Mathematics	651–670	10	604–650	49	562–603	35	500–561	6
Writing	631–670	11	593–630	51	556–592	34	500–555	4
Science	640–670	13	600–639	45	558–599	35	500–557	7
Social Science	640–670	18	599–639	48	557–598	28	500–556	6
Final Score Ranges on PSAE scale (120–200) and percent scores in each level (sample size = all tested students)								
Subject	Exceeds Standards		Meets Standards		Below Standards		Academic Warning	
	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)	Score Range	Scores within Level (%)
Reading	178–200	12	155–177	46	135–154	34	120–134	8
Mathematics	179–200	9	156–178	45	136–155	37	120–135	9
Writing	179–200	9	155–178	50	133–154	35	120–132	6
Science	178–200	11	158–177	39	136–157	38	120–135	12
Social Science	174–200	15	154–173	43	137–153	33	120–136	9