Agenda Topic: Physical Development/Health and Fine Arts Cut Scores

Materials: Report from Standard Setting

Staff Contacts: Carmen Chapman Pfeiffer
Merv Brennan

Purpose of the Agency Item:
• To inform the State Board on the cut-score recommendations for H & PD and Fine Arts

Expected Outcome of Agenda Item:
Approve H & PD and Fine Arts cut-scores for school and district information.

Background Information:

This report summarizes the approach used to establish performance categories for the ISAT fine arts and physical development/health tests at each of the three grade levels at which these ISAT tests are administered.

The development of useful and defensible standards and the execution of the standard-setting process itself are complex tasks. The State Board of Education relied on the contributions of many talented educators throughout Illinois to successfully accomplish the standard-setting process.

Prior to the meetings of the standard-setting panels themselves, ISBE convened committees of curriculum experts to develop descriptions of student knowledge and skill levels that define the four performance categories: Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. Educators throughout Illinois reviewed these knowledge and skill descriptions or definitions before sending them to the standard-setting panelists.

Panels of recognized subject matter experts subsequently convened in Champaign to translate the verbal definitions into cut scores on the ISAT tests (i.e., scores that define the boundaries between categories). Panelists were drawn from a pool of educators who had specific knowledge of student performance at the grade levels being assessed by ISAT and experience in
assessing students at those grade levels. Panelists were selected to be broadly representative of the diversity of the Illinois public school system. A total of 36 educators participated in the standard-setting process. The distribution across learning areas was as follows: physical development/health—22; fine arts—14. Each panel met for two days.

A procedure originally proposed by Angoff is one of the most frequently used methods for determining cut scores when multiple-choice test scores are used. It can be most simply described as a focused, judgmental process by knowledgeable content experts.

In the most frequent application of the Angoff method (e.g., to establish a pass/fail standard), panelists are asked to examine an item and decide what proportion of minimally competent individuals will answer the question correctly. With respect to the ISAT, however, instead of being asked about minimally competent students, panelists were asked to indicate what percentage of three groups of students—those who were just above the Academic Warning/Below Standards boundary, those who were just above the Below Standards/Meets Standards boundary, and those who were just above the Meets Standards/Exceeds Standards boundary—will answer the question correctly. The ratings were made sequentially rather than simultaneously (i.e., panelists judged the proportion of correct responses by a criterion group to every item before moving on to the next criterion group). Item performance statistics were provided to help panelists anchor their ratings. At the beginning of day two, panelists received impact data based on their day one ratings. They subsequently made a second set of ratings.

A number of checks were made on the adequacy of the ratings. Agreement among the panelists was high at all grade levels. Evaluation forms completed at the end of the session indicated that the overall level of panelist confidence in the final ratings was extremely high.

These cutoffs represent a set of fixed benchmarks against which schools can measure the success of their improvement efforts. The percentages of students who fall into each category may shift each year in response to changes in the student populations tested. However, as school improvement efforts are effectively implemented throughout the state, the expectation is that these percentages will systematically spiral upward.

Table 1 presents the results of the standard setting panels. Table 2 presents the percent of scores in each level with the proposed cut scores.
Results Table 1
ISAT Cutoffs for Each Level

<table>
<thead>
<tr>
<th>Grade</th>
<th>Physical Development/Health</th>
<th>Fine Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Academic Warning</td>
<td>Below Standards</td>
</tr>
<tr>
<td>04</td>
<td>120-137</td>
<td>138-155</td>
</tr>
<tr>
<td>07</td>
<td>120-139</td>
<td>140-155</td>
</tr>
<tr>
<td>09</td>
<td>120-131</td>
<td>132-156</td>
</tr>
</tbody>
</table>

Table 2
Percentages of Illinois Students at Each Grade Level Who Fall Into Each Category

<table>
<thead>
<tr>
<th>Grade</th>
<th>Physical Development/Health</th>
<th>Fine Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Academic Warning</td>
<td>Below Standards</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>29</td>
</tr>
</tbody>
</table>

Analysis and Implications for Policy, Budget, Legislative Action and Communications.

The PD&H and fine arts scores are for school and district information and are not used for school accountability purposes.
The recent standard setting in PD/Health and Fine Arts was based on 40 to 50 test items per grade and per area. These numbers are not much smaller than those for the other ISAT tests. Hence, the locations of the cutoffs for the four standard levels are of similar reliability.

Although standard setting used over 40 items, the actual Fine Arts and PD/Health tests are divided into forms, each of which has only 10 items on these subjects. Naturally, this small number of items does not yield a reliable measures at the individual level (the average reliability of the Fine Arts and PD/Health testlets is about 0.55 ). Therefore, the assignment of individual students to each of the four standard levels is not as reliable as for the other ISAT areas (with most reliability indices in excess of 0.91). Accordingly, it is not advisable to give out individual students scores or classifications.

Despite the low reliability of individual students' scores, the distribution of students over the four categories will stabilize at the school and district level for statistical reasons. To be sure, many individuals will fall in the wrong category. However, deviations from the "true" category frequencies will decrease as the sample size (i.e., school and district size) increases. Thus, the school and district distributions are more meaningful than the scores and the categorizations of individual students. Nevertheless, small differences between two distributions of students across the performance levels (i.e. across schools or over time) should not be over-interpreted.

Given the above, the following footnotes will be added to all school and district reports in PD/Health and Fine Arts:

- Student scores reflect only cognitive (i.e., non-performance) portions of the Illinois Learning Standards.
- For both subjects students were tested with a small number of questions. Hence, the percent in each performance level is approximate.

**Superintendent’s Recommendations:**

The Superintendent recommends the approval of the cut-scores as recommended by the Standard Setting panels.