WHAT’S GOOD ENOUGH?
Setting Benchmarks/Standards

Assessment Office, University of Hawai‘i at Mānoa  April 2012

Mission: Improve student learning through program assessment
### What’s good enough?

| “55” on the American Government College Level Examination Program (CLEP) exam | 3 credits POLS 130 |
| “3” score on the Chemistry Advanced Placement Exam | 3 credits CHEM 151 |
| 76% of students scored 5 or higher on the written communication program assessment | 76% exit ENG 100 proficient in college writing. Target met. |
Today’s Outcomes

• Understand the purpose of standard setting
• Become familiar with standard-setting methods

Agenda

1. Terminology
2. Purpose of Standard Setting
3. Key Concepts
   • Informed Judgment
   • Borderline Student
4. Multiple-choice test
5. Constructed-response test/paper/project
6. Q&A
7. Mahalo!
Terminology

OUTCOMES
Specify what students are expected to know, to be able to do, and to value

Content standards
Objectives

Terminology

STANDARDS
Define the level of performance students are expected to obtain on the outcome

Performance standards
Cutoff scores
Performance levels
Proficiency levels
Example

Novice
Proficient
Advanced

See handout, *Math Standards*

Why set standards?
Scores by themselves have no meaning.
Why set standards?

• Clarify performance expectations
• Motivate greater levels of achievement
• Allocate resources
• Decide whether to award a certificate, license, college credit
Informed Judgment

Standard setting is a blend of judgment, psychometrics, practicality.
Borderline Student

*Imagine students who are “just good enough” to meet a standard*

Just-qualified student
Minimally competent practitioner

Borderline Students

- Proficient level, 125-200
- Borderline
Borderline Students

Consult Existing Resources

- Disciplinary association standards
- Accreditation standards
- Peers' program standards
- Employers
MULTIPLE-CHOICE EXAM
Standard-setting Method

Multiple-choice Exam Standard-setting (1/2)

1. Panel of faculty + stakeholders
2. Each panelist independently estimates the probability that a borderline student will answer each test item correctly
3. Each panelist’s probabilities are summed to arrive at a panelist standard
Multiple-choice Exam Standard-setting (2/2)

4. Share and discuss
   - Actual results may be shown to give panelists an indication of consequences of particular cut scores*

5. Panelists may revise probabilities

6. Panelists’ summed probabilities are averaged for the final standard (“cut score”)

*If student results are available

CONSTRUCTED-RESPONSE: PAPER, PROJECT, ORAL, ETC.

Standard-setting Methods
Rubric-based Method (1/2)

1. Panel of faculty + stakeholders
2. Panelists discuss expectations for borderline student
3. Estimate the number of borderline students out of 100 borderline students who would receive each score possible on the rubric
4. Calculate weighted mean for each panelist.

Rubric-based Method (2/2)

5. Share and discuss
   • Actual student results may be shared to give panelists indication of consequences of particular cut scores*
6. Discuss. Panelists may revise
7. Average the panelists’ weighted means for the final standard (“cut score”)

* If student results are available
**Work-based Method (1/2)**

Needed: already-scored student work

1. Panel of faculty + stakeholders
2. Panelists review samples of student work for each score point on the rubric (without knowing the scores)
3. Panelists select the student work they view as representing the borderline student (“just good enough”)

**Work-based Method (2/2)**

4. Discuss
5. Panelists may revise their selection
6. Facilitators average the scores of the work selected as borderline to obtain the standard
Set Target for Program Success

- Examples:
  - 95% will score at the “proficient” or “advanced” level
  - 100% will meet the performance standard on doctor-patient ethics
Locally-developed Instruments

• Program’s scores/results from previous years serve as comparison or baseline

Program Results 2010 & 2011

<table>
<thead>
<tr>
<th>YEAR</th>
<th>% proficient</th>
<th>% advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>90%</td>
<td>4%</td>
</tr>
<tr>
<td>2011</td>
<td>93%</td>
<td>3%</td>
</tr>
<tr>
<td>2012 target</td>
<td>95%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Licensure Programs

• National pass rate
  • Our students’ pass rate will match or exceed the national pass rate for public institutions

• Professional accreditation requirements
  • Law school: 75% pass-rate on the bar exam
Standardized Tests

• National comparison
  • Use results from UHM peer & benchmark institutions if possible
  • Social Work program BEAP target: Our students will exceed the Public BSW-only Program mean at the 80th percentile.

Mahalo!

Assessment Office
University of Hawai‘i at Mānoa

Monica Stitt-Bergh
bergh@hawaii.edu

Marlene Lowe
mplowe@hawaii.edu
STANDARDS-SETTING WORKSHOP

Math Standards: Performance Category Descriptors & Score Ranges

*Needs Improvement (score 299 or lower)*
A student at this level
a) demonstrates partial understanding of the numeration system;
b) performs some calculations and estimations;
c) identifies examples of basic math concepts;
d) reads and constructs graphs, tables, and charts;
e) applies learned procedures to solve routine problems;
f) applies some reasoning methods to solve simple problems.

*Proficient (score 300-323)*
A student at this level
a) demonstrates solid understanding of the numeration system;
b) performs most calculations and estimations;
c) defines concepts and generates examples and counter examples of concepts;
d) represents data and mathematical relationships in multiple forms (e.g., equations, graphs);
e) applies learned procedures and mathematical concepts to solve a variety of problems, including multi-step problems;
f) uses a variety of reasoning methods to solve problems and explains steps and procedures.

*Advanced (score 324 or higher)*
A student at this level
a) connects concepts from various areas of mathematics, and uses concepts to develop generalizations;
b) performs complex calculations and estimations;
c) selects the best representation for a given set of data and purpose;
d) generates unique strategies and procedures to solve non-routine problems;
e) uses multiple reasoning methods to solve complex problems, and justifies strategies and solutions.

STANDARDS-SETTING WORKSHOP
MULTIPLE-CHOICE EXAM STANDARD-SETTING ACTIVITY

World War II History Exam Excerpt: College-entry Level

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>What is the probability that a borderline proficient student will answer correctly?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. President Harry Truman’s decision to use atomic bombs against Japan was primarily based on his belief that</td>
<td></td>
</tr>
<tr>
<td>A. an invasion of Japan would result in excessive casualties</td>
<td></td>
</tr>
<tr>
<td>B. Germany would refuse to surrender in Europe</td>
<td></td>
</tr>
<tr>
<td>C. an alliance was developing between Japan and the Soviet Union</td>
<td></td>
</tr>
<tr>
<td>D. Japan was in the process of developing its own atomic weapons</td>
<td></td>
</tr>
<tr>
<td>2. Which action best illustrates the policy of isolationism followed by the United States before it entered World War II?</td>
<td></td>
</tr>
<tr>
<td>A. signing of a collective security pact with Latin American nations</td>
<td></td>
</tr>
<tr>
<td>B. passage of neutrality legislation forbidding arms sales to warring nations</td>
<td></td>
</tr>
<tr>
<td>C. embargo on the sale of gasoline and steel to Japan</td>
<td></td>
</tr>
<tr>
<td>D. President Franklin D. Roosevelt’s exchange of American destroyers for British naval and air bases</td>
<td></td>
</tr>
<tr>
<td>3. Which precedent was established by the Nuremberg war crimes trials?</td>
<td></td>
</tr>
<tr>
<td>A. National leaders can be held responsible for crimes against humanity.</td>
<td></td>
</tr>
<tr>
<td>B. Only individuals who actually commit murder during a war can be guilty of a crime.</td>
<td></td>
</tr>
<tr>
<td>C. Defeated nations cannot be forced to pay reparations.</td>
<td></td>
</tr>
<tr>
<td>D. Defeated nations can be occupied by the victors.</td>
<td></td>
</tr>
<tr>
<td>4. “The great rule of conduct for us in regard to foreign nations is, in extending our commercial relations, to have with them as little political connection as possible.”</td>
<td></td>
</tr>
<tr>
<td>This quotation supports a foreign policy of</td>
<td></td>
</tr>
<tr>
<td>A. imperialism</td>
<td></td>
</tr>
<tr>
<td>B. appeasement</td>
<td></td>
</tr>
<tr>
<td>C. neutrality</td>
<td></td>
</tr>
<tr>
<td>D. economic sanctions</td>
<td></td>
</tr>
</tbody>
</table>


1) A 2) B 3) A 4) C
What is the probability that a \textbf{borderline proficient student} will answer correctly?

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>Panelist #1</th>
<th>Panelist #2</th>
<th>Panelist #3</th>
<th>Panelist #4</th>
<th>Panelist #5</th>
<th>Panelist #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. President Harry Truman’s decision to use atomic bombs against Japan was primarily based on his belief that</td>
<td>0.8</td>
<td>0.25</td>
<td>0.3</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>2. Which action best illustrates the policy of isolationism followed by the United States before it entered World War II?</td>
<td>0.8</td>
<td>0.25</td>
<td>0.3</td>
<td>0.2</td>
<td>0.8</td>
<td>0.25</td>
</tr>
<tr>
<td>3. Which precedent was established by the Nuremberg war crimes trials?</td>
<td>0.3</td>
<td>0.25</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>4. &quot;The great rule of conduct for us in regard to foreign nations is, in extending our commercial relations, to have with them as little political connection as possible.&quot; This quotation supports a foreign policy of</td>
<td>0.7</td>
<td>0.25</td>
<td>0.5</td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**SUM OF PANELIST'S PROBABILITIES**  
2.6 1 1.5 1.7 2.7 1.55

**AVERAGE OF PANELISTS' PROBABILITIES**  
1.8

\textit{Students who score below are not proficient.}  
\textit{Students who score at or above or higher are proficient.}
# Standards-Setting Workshop

## Constructed-Response (Paper, Project, Oral, etc.) Standards-Setting Activity

Estimate the number of students **out of 100 borderline proficient students** who would fall into each score category (total must be 100)

<table>
<thead>
<tr>
<th>Oral Communication Rubric</th>
<th># of borderline proficient students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is not observable within the presentation. Language choices are unclear and minimally support the effectiveness of the presentation. Language in presentation may not be appropriate to audience. Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) detract from the understandability of the presentation, and speaker appears uncomfortable. Insufficient supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) make reference to information or analysis which minimally supports the presentation or establishes the presenter's credibility/authority on the topic. Central message can be deduced, but is not explicitly stated in the presentation.</td>
<td></td>
</tr>
<tr>
<td><strong>2</strong> Organizational pattern is intermittently observable within the presentation. Language choices are mundane and commonplace and partially support the effectiveness of the presentation. Language in presentation is appropriate to audience. Delivery techniques make the presentation understandable, and speaker appears tentative. Supporting materials make the presentation understandable, and speaker appears tentative. Supporting materials make appropriate reference to information or analysis which partially supports the presentation or establishes the presenter's credibility/authority on the topic. Central message is basically understandable but is not often repeated and is not memorizable.</td>
<td></td>
</tr>
<tr>
<td><strong>3</strong> Organizational pattern is clearly and consistently observable within the presentation. Language choices are thoughtful and generally support the effectiveness of the presentation. Language in presentation is appropriate to audience. Delivery techniques make the presentation interesting, and speaker appears comfortable. Supporting materials make appropriate reference to information or analysis which generally supports the presentation or establishes the presenter's credibility/authority on the topic. Central message is clear and consistent with the supporting material.</td>
<td></td>
</tr>
<tr>
<td><strong>4</strong> Organizational pattern is clearly and consistently observable and is skillful and makes the content of the presentation cohesive. Language choices are imaginative, memorable and compelling and enhance the effectiveness of the presentation. Language in presentation is appropriate to audience. Delivery techniques make the presentation compelling, and speaker appears polished and confident. A variety of types of supporting materials make appropriate reference to information or analysis which significantly supports the presentation or establishes the presenter's credibility/authority on the topic. Central message is compelling: precisely stated, appropriately repeated, memorable, and strongly supported.</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from the Oral Communication VALUE Rubric (Association of American Schools & Colleges)

---

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.
Estimate the number of borderline proficient students who would receive each score (total = 100 students)

<table>
<thead>
<tr>
<th>RUBRIC SCORE</th>
<th>Panelist #1</th>
<th>Panelist #2</th>
<th>Panelist #3</th>
<th>Panelist #4</th>
<th>Panelist #5</th>
<th>Panelist #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>5</td>
<td>40</td>
<td>25</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>13</td>
<td>40</td>
<td>70</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>80</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>25</td>
</tr>
</tbody>
</table>

Subtotal # of students: 100 100 100 100 100 100

Weighted Average: 2.3 3.7 2.5 2.8 2.6 3.0

**Average of Panelists' Weighted Avg.**: 2.8

Cut off score = 2.8

*Students who score below are not proficient.*

*Students who score at or above are proficient.*