

Setting and Reviewing the NPTE Passing Standard

September 30, 2005

Mary J. Pitoniak, PhD
Cynthia A. Searcy, PhD

1

Note:

*Portions of this presentation were given
at the 2005 FSBPT Annual Meeting
in Austin, Texas
by Gregory J. Cizek, PhD and
Mary J. Pitoniak, PhD*

2

Standard Setting

- Standard setting refers to the process of establishing a cut score—a score that divides the score range into passing and failing.
- Examinees at or above the cut score pass, and those below it fail.
- “The cut score represents an informed judgment that those scoring below it are likely to make serious errors for want of the knowledge or skills tested” (AERA/APA/NCME, *Standards for Educational and Psychological Testing*, 1999, p. 53).

3

Standard Setting (continued)

- “No test is perfect, of course, and regardless of the cut score chosen, some examinees with inadequate skills are likely to pass and some with adequate skills are likely to fail.”
- “Exposing the public to potential harm by issuing a license to an incompetent individual (false positive) must be weighed against . . . denying a license to, and thereby disenfranchising, a qualified examinee (false negative).”

(AERA/APA/NCME, 1999, p. 54)

4

Importance of the Standard Setting Process

“Where the results of the standard-setting process have highly significant consequences . . . those responsible for establishing the cut scores should be concerned that the process by which cut scores are determined be clearly documented and defensible.” (AERA/APA/NCME, 1999, p. 54)

5

Important Caveats

- No “true” value exists
- Policy decision, related to mission of organization (e.g., public protection)
- Must be revisited when test specifications, scope of practice, preparation programs, other factors change

6

Setting Performance Standards

1. Choose Method
2. Select Qualified Participants
3. Train Participants
4. Provide Feedback
5. Calculate the Cut Score
6. Gather Validity Evidence

7

Choose Method

“When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of item or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge to bear in a reasonable way.”

(AERA/APA/NCME, 1999, p. 60)

8

Method (continued)

- Modified Angoff (1971) method
 - criterion-referenced
 - medical and health-related credentialling
 - vast research base
 - defensible

9

Method (continued)

- Standard setting conducted in September 2002
- Context: Item compromise, concern about validity of test scores; test form withdrawn
- Done on whole test forms, separately for each test form
- Required use of some items lacking performance data; lacked impact data

10

Select Qualified Participants

- Representative of practice settings and specialties, regions
- n = 12
- Knowledgeable regarding practice analysis, test specifications

11

Train Participants

- "Minimally qualified candidate"
 - key referent to be operationalized in Angoff method
 - borderline performance
- "Minimal knowledge, judgment, technical and interpersonal skills required to safely practice physical therapy, including examination, evaluation, diagnosis, prognosis, interventions, and outcomes assessment."

12

Provide Feedback

- "Informed Consensus"
- Rationales for high, low ratings
- Referral to key referent

13

Calculate the Cut Score

Hypothetical Round 1 Data

	I1	I2	I3	I4	I5	Means
Raters 1	60	55	80	85	70	70
2	60	50	75	80	75	68
3	55	55	80	85	75	70
4	85	95	90	100	85	91
5	20	15	35	50	30	30
Means	56	54	72	80	67	65.8

14

Calculate the Cut Score (continued)

- 65.8 (percentage of items that minimally competent candidate would be expected to answer correctly)
- Multiply by number of operational (scored) items in test form
- Consider relative costs of false positives, false negatives

15

Results from September 2002

- Defensible
- Easily explained
- Well-implemented
- Continual validation

16

Gathering Validity Evidence

- Pass Rate Trends
- Two Audits
 - Commission
 - HumRRO
- Input from employers, educators, states
 - Some feedback that cut score was too high
 - Some feedback that cut score was too low
- Concluded that cut scores were sound, but that review was appropriate

17

Review of Passing Standard

- Standard review session conducted April 22, 2005
- **Direct Consensus Method** used
- In general, same five steps as Angoff method were employed, though the details of their implementation differed

18

Panelists

- Solicited nominations from APTA Section Chairs and Jurisdiction Board Members and Administrators
- Received 275 nominations (269 with contact information)
- 88 nominees applied and were available on April 22, 2005 for the panel

19

Panelists

- 12 panelists
- Representative and knowledgeable, as with previous standard setting
- Range of years of service—
 - Minimum: 2
 - Maximum: 30
 - Median: 19

20

Other Panelist Characteristics

Characteristic	Category	#
Gender	Female	8
	Male	4
Ethnicity (observed)	Black	2
	White	10
Region	Northeast	3
	Southeast	2
	North Midwest	1
	South Midwest	2
	Rocky Mountain	2
	Southwest	1
	Northwest	1

21

Characteristic	Category	#
Specialty	Acute care	1
	Administration	2
	Cardiovascular/pulmonary	1
	Geriatrics	1
	Neurology	2
	Orthopedics	2
	Pediatrics	2
	Sports medicine	1
	Setting	Academics: Clinical faculty
Academics: Faculty		2
Academics: Program administrator		1
Geriatrics		1
Home health		1
Hospital/acute care		2
Outpatient facility (orthopedic)		1
Pediatric (school or pediatric clinic)		1
Private practice		1
Rehabilitation/academics		1

22

Observers

- To make the process as transparent as possible, a group of 9 physical therapists who were not selected to be on this panel was invited to observe.
- They viewed the process via closed-circuit TV and reviewed the test materials that the panelists used in order to see exactly how the process was conducted.
- They did not, however, interact with the panelists, so as not to influence the process.

23

Direct Consensus Method

- Panelists review clusters of items that are **grouped by content**
- Then they provide a **rating** for each cluster:

How many items would the minimally competent examinee answer correctly?

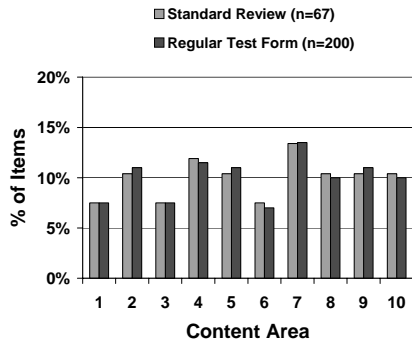
24

Content Areas

1. History and Systems Review
2. Test and Measures Group I
3. Test and Measures Group II
4. Evaluation and Diagnosis
5. Prognosis and Outcomes
6. Non-Procedural Intervention
7. Procedural Intervention Group I
8. Procedural Intervention Group II
9. Procedural Intervention Group III
10. Standards of Care

25

Distribution of Items



Sample Rating Form

Cluster	# of Items in Cluster	Estimated # Correct
1	5	4
2	7	6
3	5	3.5

Note: Data are hypothetical.

27

Sample Spreadsheet Display

Panelist	Cluster			Passing Score
	1	2	3	
1	3	4	3	32
2	4	7	3	35.5
3	3.5	6	4.5	34
4	4	5	3	38
5	4	5	3.5	37.5
12	4	4.5	4	33

Note: Data are hypothetical.

28

After Ratings

- Panelists discuss ratings
- Feedback given:
 - ❖ How minimally competent examinees performed on items
 - ❖ How many examinees would pass given that passing score
- Panelists given opportunity to revise ratings
- Panelists encouraged to come to consensus

29

Evaluation of Meeting

- Did panelists appear to understand the rating task?
- Did feedback provided to panelists appear to have an effect?
- Were panelists positive about their ratings and the result of the meeting?

30

Panelist Evaluations

- It is important to look at panelists' opinions after training as well as at the end of the meeting.
- In this study, panelists did appear to understand the task and feel confident in their ratings and their results.

31

Sample Panelist Evaluation Results

- How confident are you with your understanding of the description of the minimally competent physical therapist?

	Before Training	After Training
❖ Very Clear	4	7
❖ Clear	6	5
❖ Somewhat Clear	2	0

32

Panelist Evaluations (continued)

- What level of confidence do you have that the panel has recommended an appropriate and defensible passing standard?

❖ Very Confident	4
❖ Confident	6
❖ Somewhat Confident	2

33

Panelist Evaluations (continued)

- What is your opinion of the passing standard recommended by your panel?

❖ About Right	9
❖ Too High	1
❖ Too Low	2

34

Observer Feedback

- All of the observers felt the process and methodology were well thought-out and understandable.
- In general, observers felt the panel came to an agreement on the definition of minimal competence, though several felt that there were a few panelists that didn't agree with the rest of the group.
- All of the observers felt that having a group of observers was very important and that they saw their role as being "ambassadors for the process."

35

Result of Meeting

- The panelists did not come to a consensus on the passing standard.
- Therefore, the average of the panelists' passing standards was used, as specified by the Direct Consensus Method procedures.
- That average passing score was **lower** than the one in place up until that time, as will be reviewed next.

36

Steps Taken After Standard Review Meeting

- Review of results by legal counsel, FSBPT staff, technical advisory panel
- FSBPT Board meeting on May 12, 2005 to adopt new standards
- Immediate implementation

37

Frequently Asked Questions

- By what means did the FSBPT attempt to notify stakeholders (candidates, State Boards, faculty) of its intent to review the standard?
 - Notice on FSBPT public and faculty websites, April 1, 2005
 - Article in *Newsbrief*, April 11, 2005
 - Update to notice on FSBPT public and faculty websites, April 28, 2005
 - *Newsflash* (and subsequent posting on FSBPT public and faculty websites), May 12, 2005
 - FAQ communications with Board Administrators for all U.S. jurisdictions (and subsequent posting on FSBPT public and faculty websites), May 13, 2005
 - Article in *Quarterly Faculty Newsletter*, June, 2005
 - Article in *Forum*, July, 2005

38

Frequently Asked Questions (continued)

- How much easier is the cut score?
 - Raw score difference varies from one form to another
 - Expect the pass rate to stabilize to about 80% to 86%
- What is the current pass rate?
 - Approximately 10% to 12% higher than this time last year

39

Frequently Asked Questions

(continued)

- Why was the new cut score implemented on May 12, 2005?
 - Procedures to establish cut score in 2002 were evaluated and found to be sound
 - 2005 forms introduced before FSBPT Board decided to review cut score
 - No psychometric or legal justification exists for retrospective scoring, according to the Technical Advisory Panel and legal counsel
 - Postponing implementation until introduction of 2006 forms was deemed too long to wait

40

■ Thank you.

■ Questions?

41
