Standard Setting in Health Professions Education

Shital Bhandary

Department of Community Medicine
Patan Academy of Health Sciences, Lalitpur, Nepal

'Standard setting' is a process through which the items prepared for the student assessment is standardized before and/or after the student assessment process. The main objective of standard setting is to determine the cut-off score, or pass mark, of an item or a test. The tests represent Multiple Choice Questions (MCQs), Short Answer Questions (SAQs), Objective Structured Practical or Clinical Examination (OSPE/OSCE) as well as Problem Based Learning (PBL) and similar checklist based methods.[1]

There are different standard setting methods practiced around the world.[1,2]

Fixed or Holistic Standard Setting
Norm-Referenced Standard Setting
Criterion-Referenced Standard Setting

In the fixed or holistic standard setting method, the cut-off score is determined *a priori* by the university and even the regulatory bodies.[2,3] The most common or preferred pass mark in health professions' education is 50 percent of the total marks and is set before the assessment. However, the question is, can we be satisfied with health care providers who are competent in only 50 percent of the knowledge and skills laid down in the curriculum? The general public, who will receive health care from these half competent health care professionals, would not be happy to know this fact. Even if the fixed standard is set at higher level of say 60 percent to 80 percent, it is very difficult to defend such cut-off scores as it is based on arbitrary numbers.

The norm-referenced standard setting, on the other hand, has evolved as the alternative to the fixed or holistic standard setting method.[3,4] Here, the cut-off score is determined using statistical theories and methods. The most preferred one is declaring anyone who scores less than minus two standard deviation (- 2 SD) from mean as 'failed' or 'incompetent'. This method is applied after the test and standardized for the particular cohort of students that take the test. There are practices of using percentile, Z-score, T-score, Stanine scores as well. The main limitation of this method is that it is based on the performance of the cohort of student e.g. same batch, year, semester etc. in a test, which will be different from cohort to cohort and results are not comparable across cohorts based on their performance. Another limitation of this method is that test scores always do not follow the Gaussian or Normal distribution and the notion of eliminating those with bottom 2.5% of the test sore does not follow well with the underlying assumptions on which it is based.

To compensate the fallacies of the fixed and norm referenced standard setting methods, criterion-reference standard setting methods have been formulated.[1,2,3,4] The standard setting of MCQ, for instance, is based on the assumption that test items should be evaluated by at least six subject matter experts to ensure content and construct (difficulty levels) validity and reliability. In addition, the standard setting group should contain at least two non-content experts to ensure face validity of the item and to minimize the content expert's scoring bias (i.e. higher scores from the panel of content experts even for the "borderline candidates" or those who may pass or fail the test). The Minimally Competent Borderline Candidates (McBC) or simply the 'Borderline candidates' are the hypothetical cohort of students who would either pass or fail the item given the construct and content of the item.[2,5]

The borderline candidate score for Structured Short Answer Questions (SSQA) would be a range of scores where the cut-off score would fall. It is determined *a priori* by the standard setting group that is involved in writing the SSQA. In practical assessments like OSPE/OSCE as well as process assessments like PBL the notion of 'borderline candidate' are assessed as part of the global rating of the checklist where the assessor gives his or her verdict for an individual student based on his/her overall performance on the domain being assessed i.e. knowledge, skill, behavior etc.

Standard setting can be done before, at the time of, or after the test.[1,2] The most common method of standard setting that determines the pass mark before the test is Modified Angoff method where a group of at least 6 mixed judges sit down and discuss the content and construct of the item.[1,2] Once they reach consensus then each judge gives his or her cut-off mark relating it with the 'borderline candidate'.[2] So, the difficult items will get low scores whereas easy items will get high scores. Yet, items from the 'must know' areas will have higher scores irrespective of it being easy or difficulty construct. The judges' scores are then averaged and cut-off score of an item is determine and flagged as Angoff score. The cut-off score of a test with 100 such items will

be determined by taking average of their Angoff scores.[2] Thus, pass mark of a test is known *a priori*.

The Modified Angoff method can also be used to standard set SSAQs as well as checklist based assessment methods. The Angoff score will be high in the beginning and will stay high if it is not done properly i.e. violating the cut-off score being based on the 'borderline candidate'.[1,2] Thus, it is recommended to compensate the judges' bias by using compromise methods where the actual test score of the cohort and Angoff score are used in combination to determine the cut-off score. The most popular and recommended compromised methods are Hoftsee[2] and Buek.[6]

Another popular method of standard setting is "borderline method" and is usually done after the test.[1,2] Here the cut-off score is obtained from the borderline score range as in SSQA or borderline global rating as in OSPE/OSCE and PBL checklists. In SSQA a simple regression method is used to determine the cut-off score whereas median score of the borderline candidates becomes the cut-off score in the skill (e.g. OSPE/OSCE) and behavior checklist scores (e.g. PBL).

The criterion-referenced standard setting methods should be used by aligning them with the curricular objectives and proper examination blueprint. This process controls the curriculum mismatch i.e. curriculum, teaching and assessment not being matched with each other. This process also ensures the minimum competency that is required to pass a course and curriculum as per the objectives set *a priori*.

Proper sampling of the items, or blueprinting, ensures the competency in terms of knowledge, skills and behaviors that are linked to the curriculum, which in turn are linked to the institutional goals and even the mission statements of the university. The hallmark of effective student assessment is that it captures them all. Further, involvement of at least six faculty members to ensure reliability and validity of each item clearly out-weighed advantages of the items being constructed by a single faculty, no matter how much of an expert he or she is in the subject matter. This process further limits the choice of an individual faculty prevailing in the student assessment and is considered more student-centered.[1,2,3]

A health science education institute (medicine,nursing,allied health sciences etc.) with integrated teaching and learning should use integrated assessment methods and involve faculty from different subjects and specialties and, the criterion-reference standard setting ensures it is being done properly and effectively. [2-6] The main criticism of the criterion-method is the concept of 'borderline candidate' but none of the assessment methods would be complete without incorporating subjectivity within the objectivity. The teachers, facilitators and faculty involved in the teaching and learning methods can easily think of those students who "may pass or fail" based on their own experience, which then becomes the basis for defining and defending the 'borderline candidates'.

REFERENCES

1. Cizek, GJ, editor. Setting Performance Standards: Concepts, Methods and Perspectives. New Jersey: Lawrance Erlbaum Associates Inc., 2001.

2. Norcini JJ. Setting standards on educational tests. Med Educ 2003 May; 37(5): 464-9.

3. Kaufman DM, Mann KV, Muijtjens AMM, van der Vlueten CPM. A Comparison of Standard-setting Procedures for and OSCE in Undergraduate Medical Education. Academic Medicine 2000 Mar; 75 (3): 267 – 71 .

4. George S, Haque MS, Oyebode F. Standard Setting: Comparison of two methods. BMC Med Educ 2006 Sept, 6: 46.

5. Carlson J, Tomkowiak J, Stilp C. Using the Angoff Method to set Defensible CutOff Scores for Standardized Patient Performance Evaluation in PA Education. J Physician Assist Educ 2009; 20(1): 15 – 23.

6. Wendt A, Kenny L. Setting the Passing Standard for the National Council Licensure Examination for Registered Nurses. Nurse Educ 2007 May-Jun; 32(3): 104 –8