

Standard setting and quality of assessment: A conceptual approach

S S Banda, MB ChB (UNZA), MSc (Glasgow), MMed (Dundee), PhD (UNZA), FAcadMed (UK)

Department of Medical Education Development, School of Medicine, University of Zambia, Lusaka; and School of Medicine, Cavendish University, Lusaka, Zambia

Corresponding author: S S Banda (ssbanda2007@gmail.com)

The debate with regard to standard setting continues among scholars, but the perfect method remains elusive. Nonetheless, resolute demands for accountability by means of defensible, valid and reliable practices, including pass or fail decisions, provide an opportune vehicle for scholars to consider the quality of assessments and ramifications on standard setting. This conceptual review considers how standard setting can be placed strategically in the quality dialogue space to address concerns about credibility and defensibility in the literature. Quality performance standards and the effect of assessment outcomes are important in the educational milieu, as assessment remains the representative measure of achievement of the prescribed performance standard but also includes the quality of health professions education (HPE). The author suggests that standard setting can be a pivotal focus for technical and psychometric sufficiency of assessments and accountability of HPE institutions towards stakeholders. Standard setting should not be seen as a methodological process of setting pass/fail cut-off points only, but as a powerful catalyst for quality improvements in HPE by promoting excellence in assessments.

Afr J Health Professions Educ 2016;8(1):9-10. DOI:10.7196/AJHPE.2016.v8i1.712



The practice of deciding the pass/fail cut-off point is commonly referred to as standard setting.^[1] However, many standard-setting practices, e.g. the historical 50% pass/fail cut-off point, can be difficult to defend psychometrically, yet they continue to be used despite rational disapproval in the literature.^[2-7] Demands for accountability by means of defensible, valid, reliable and robust assessment policies and practices, including pass/fail decisions, are increasingly becoming important.^[5,8] Discussions are being held on credible and defensible standard-setting practices^[6,7] for high-stakes examinations. This debate presents a vehicle to discuss quality of assessment, which has a crucial role in promoting the quality of health professions education (HPE) institutions.^[9] This conceptual review considers how standard setting can be placed strategically within the quality dialogue arena to ensure that appropriate action is taken to address concerns about credibility and defensibility in the literature. The medical education communities in African countries are small and need to join forces to contribute to the dialogue on standard setting and its potential to improve the quality of HPE.

This article also provides the background to the purpose of assessments and sources of common standard-setting procedures. The last section gives a conceptual framework to embed standard setting in the HPE quality discourse.

Assessment purpose

Assessment categorises examinees into two groups – those who pass and those who fail – with regard to curriculum objectives, content and skills.^[5] The final mark represents an examinee's attainment on the performance continuum implied by proficiency levels^[5] and is represented on a test score range from 0% to 100%.

This process means that standard setting is the embodiment of the attainment of learning objectives^[10] and the pass mark is the operationalisation of the performance standard.^[11] A decision has to be made whether the cut-off marks and the resulting pass/fail results are sufficiently valid and accurately representative of the intended interpretations assigned to them. For specific information about the technical process and in-depth review of each method,

several publications may be consulted.^[1-3] There are ~50 standard-setting methods reported in the literature.^[7] Some of the well-known methods include the Angoff method;^[2,4,8,10] Ebel method;^[1,2,10] Bookmark method;^[12,13] Borderline-group method and contrasts by group approach.^[4,6,7,10]

A conceptual approach

Plain, fair and valid standard setting cannot result from unfair and invalid assessments.^[7] Validity, reliability, educational effect, feasibility and acceptability frequently constitute the criteria used to select assessment methods.^[14] Primarily, validity and reliability have been central to decisions about the assessment methods used in different HPE settings. Validity focuses on whether a test succeeds in testing the competencies for which it was designed.^[14] Reliability or generalisability is a measure of the relative size of variability in scores due to error, with the objective of reaching a desired level of measurement accuracy across different tasks.^[14] Acceptability is the extent to which many stakeholders endorse the measure and interpretation of scores.^[14] Blueprinting, which requires test content to be carefully planned against learning objectives,^[1,5] has also been advocated. The focus should be on psychometric rigour.^[14] Because test theory, e.g. classic test theory, item response theory, generalisability theory,^[1] informs psychometrics, it must be a critical consideration in the assessment of planning, analyses and interpretation, especially with regard to reliability. The abovementioned considerations can be grouped into 'assessment technical sufficiency' matters.

However, an important result of the standard-setting process is the pass/failure rate, which directly or indirectly implicates the quality of instruction, instructors and graduates. The reason is that assessment is the symbolic measure of confirming achievement of the prescribed performance standard. Together, these can be considered 'impact of assessment outcomes'.

Educational tension, therefore, arises between the expectations of quality of assessment, performance standards and impact of assessment outcomes.^[7] Accordingly, technical sufficiency of assessment practices is not enough – impact considerations are as important. The close relationship between quality of assessment and standard-setting practices and their effect on HPE accountabilities is neglected, poorly understood and underdeveloped at many HPE

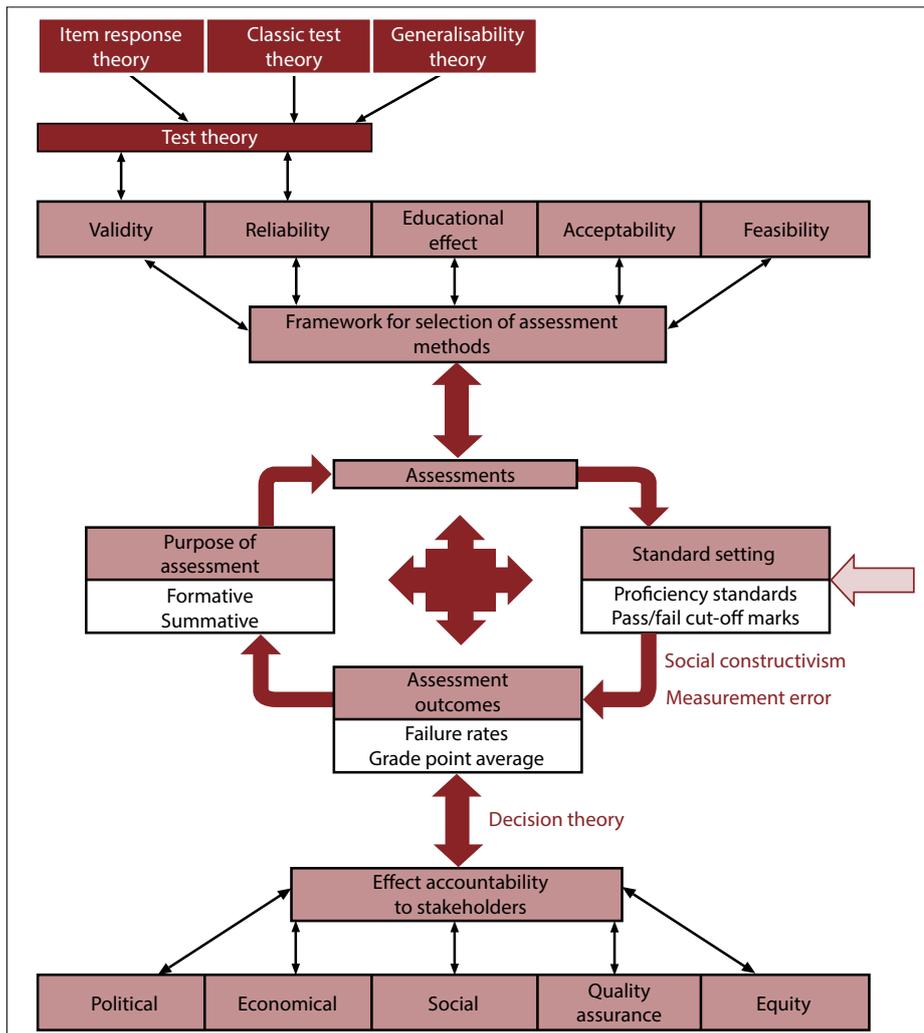


Fig. 1. Conceptual framework of standard setting as quality assurance catalyst.

institutions. The ultimate key determinants of pass rates are the quality of the assessments and the pass/fail standard-setting practices and decisions. In the context of their validity, educators must be aware of this relationship as they plan, implement and interpret assessment outcomes. Consequently, this article proposes that standard setting should be seen as a critical aspect of the assessment structures and processes. It is dependent on assessment quality, as a test that does not cover appropriate content is not at the appropriate level of difficulty, is not reliable, and will not lead to appropriate decisions, regardless of the standard-setting methods employed.^[11] However, educational assessments

not only comprise technical sufficiency; attention must also be paid to the accountability of the HPE institution to various stakeholders. The holistic representation is captured in the conceptual approach (Fig. 1), summarising key principles and concepts in a relational method.

Standard setting can be a pivotal focus to attend to technical and psychometric sufficiency of assessments as well as accountability responsibilities of HPE institutions to stakeholders. Although standard setting is an important psychometric problem, it is not solely a technical issue.^[1,11] The consequences of appropriate or inappropriate standards for society, institutions and individuals must be considered.^[11]

Crocker and Zieky^[7] noted that because of the judgemental nature of standards, validity was dependent on how sensible the standard-setting process and its outcome were, together with the consequences of pass rates and possible classification errors. The effect of assessment outcomes therefore extends to political, economic, social and policy domains.^[1,7,11] Notably, decisions about the effect of assessments on accountability have to depend on the decision-making theory^[7] or alternative decision theories.

Conclusion

Standard setting should not be seen as a methodological process of setting the pass/fail cut-off point only, but as a powerful catalyst for quality improvement in HPE by promoting assessment excellence. While medical education departments can be important catalysts with regard to capacity and quality, in Africa, for example, their development remains basic. Therefore, numerous academic staff and policy makers should join the standard-setting dialogue and in particular its potential to improve the quality of HPE.

References

1. Crocker L, Algina J. Introduction to Classical and Modern Test Theory. Boston, MS: Wadsworth, 2006.
2. Cizek GJ. Setting Performance Standards: Foundations, Methods, and Innovations. London, UK: Routledge, 2012.
3. Cizek GJ, Bunch MB. Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests. Thousand Oaks, CA: Sage, 2007.
4. Ben-David MF. Standard setting in student assessment. *Med Teach* 2000;22(2):120-130.
5. Hambleton RK. Setting performance standards on educational assessments and criteria for evaluating process. In: Cizek GJ, ed. *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum, 2001:89-116.
6. Zieky MJ, Piere M, Livingston S. Cut-scores: A manual for setting standards of performance on educational and occupational tests. 2008. <http://www.amazon.com/cut-scores-standards-performance-educational-occupational/dp/1438250304> (accessed 4 July 2015).
7. Crocker L, Zieky M. Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessment, Washington, DC, 5 - 7 October 1994. Washington, DC: National Center for Education Statistics, 1995.
8. Kilminster S, Roberts T. Standard setting for OSCEs: Trial of borderline approach. *Adv Health Sci Educ Theory Pract* 2004;9:201-209.
9. Barman A. Standard setting in student assessment: Is a defensible method yet to come? *Ann Acad Med Singapore* 2008;37:957-963.
10. Ricker K. Setting cut scores: Critical review of Angoff and modified-Angoff methods. *Alberta Journal of Educational Research* 2006;52(1):53-64.
11. Norcini JJ, Banda SS. Increasing the quality and capacity of education: The challenge for 21st century. *Med Educ* 2011;45:81-86. [<http://dx.doi.org/10.1111/j.1365-2923.2010.03738.x>]
12. Karantonis A, Sireci SG. The Bookmark standard setting method: A literature review. *Educational Measurement: Issues and Practice* 2006;4:12.
13. Buckendahl C, Smith R, Impara J, Plake B. A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational Measurement* 2002;39(3):253-263.
14. Norcini JJ, McKinley DW. Assessment methods in medical education. *Teaching and Teacher Education* 2007;23:239-250.