

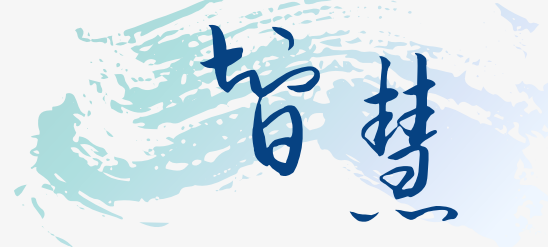
Finding the standard setting method for your assessment

<https://doi.org/10.29060/TAPS.2025-10-3/TT003>

Standard setting is the process of deciding the boundary or standard that separates the candidates into two (e.g. pass and fail) or more groups, based on their ability shown at an assessment. Standard setting methods can be broadly grouped into four clusters (see table below).

When to use which method, though a crucial decision for any Board of Examiners, is inadequately explored in the literature. The following brief guide attempts to bridge this literature gap.

Cluster of Methods	Key Features	Issues	When to Use
Arbitrary standards and norm-referenced standards	<ul style="list-style-type: none">Arbitrary standards produce a fixed pass mark, e.g. candidates scoring 50% or more pass.Norm-referenced standards produce a fixed pass rate, e.g. 40% of top-scoring candidates pass.	The pass mark is unrelated to the difficulty of assessment items.	<ul style="list-style-type: none">Arbitrary standards: not indicated for high-stakes assessment.Norm-referencing: used for selection purposes.
Test-centred methods	<ul style="list-style-type: none">A group of experts (judges) estimate the probability of a hypothetical borderline (a candidate who has a 50% probability of passing or failing) or a just-passing candidate passing the test items, e.g. Angoff (1971), Ebel (1972), Nedelsky (1954), Bookmark (Karantonis & Sireci, 2006), Jaeger (1982).The judges' estimates are collated through an averaging process.An expert (judge) is a subject-matter specialist, with considerable experience as a teacher and an assessor, well versed with the educational basis behind standard setting.	<p>Although the pass mark is directly related to the difficulty of test items,</p> <ul style="list-style-type: none">human judgement is not infallible: The pass mark can vary from one panel of judges to another, even for the same test.finding a sizeable group of experts (at least 8) satisfying all requirements is difficult.it is difficult for judges to visualise a hypothetical borderline candidate.the process is time-consuming. <p>Due to the above difficulties, the pass mark can be unrealistic.</p>	<ul style="list-style-type: none">When an adequate number of properly trained and experienced expert judges who can devote quality time to the standard setting process is available.When modifications such as the Modified Angoff method can be used to overcome unrealistic standards by allowing judges to be informed by actual results of previous similar exams.
Partially results-based methods-I: Examinee-centred methods	<ul style="list-style-type: none">Based on actual candidate performance, judges group candidates into two or more groups, e.g. Borderline group (Smee & Blackmore, 2001), Borderline regression (Kramer et al., 2003), Contrasting groups (p.35) (Livingston & Zieky, 1982) and Up-down (p.43) (Livingston & Zieky, 1982) methods.The pass mark is calculated using the actual candidate scores.	Although judgements are realistic, the introduction of actual test results tends to make the standard cohort-dependent, i.e., norm-referencing features influence the standard.	<ul style="list-style-type: none">When there is a sufficiently large number of candidates.When a global score or a global pass/fail decision is available, in addition to the usual itemised score.When the judges are well-trained in making a global decision independent of the itemised scores.



Cluster of Methods	Key Features	Issues	When to Use
Partially results-based methods-II: Compromise methods	<ul style="list-style-type: none"> Judges make judgements by looking at test items, and those judgements are superimposed on actual candidate scores to derive the pass mark, e.g. Hofstee method (Hofstee, 1973). 	<ul style="list-style-type: none"> Expert judgements and actual results may not match each other. The standard can be cohort-dependent due to the norm-referencing features of actual candidate scores. 	<ul style="list-style-type: none"> When trained judges, actual results of a sizable cohort of candidates and expertise in handling both judges' judgements and results are available. Mostly used as a backup method to verify standards generated by other methods.
Results-based methods	<ul style="list-style-type: none"> Judges are not needed for standard setting. The pass mark is generated by statistically manipulating the actual marks, e.g. Cohen (Cohen-Schotanus & van der Vleuten, 2010) and Wijnen (1971) methods. 	Due to the norm-referencing influence, the pass mark could be high and defensibility would be an issue.	These methods should be used in high-stakes assessment only when an adequate evidence base is built by conducting them parallelly with another more established method.

References

- Angoff, W. H. (1971). *Scales, norms, and equivalent scores*. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). American Council on Education.
- Ebel, R. L. (1972). *Essentials of educational measurement*. Prentice Hall.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14(1), 3-19. <https://doi.org/10.1177/001316445401400101>
- Karantonis, A., & Sireci, S. G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement Issues and Practice*, 25(1), 4-12. <https://doi.org/10.1111/j.1745-3992.2006.00047.x>
- Jaeger, R. M. (1982). An iterative structured judgment process for establishing standards on competency test: Theory and application. *Educational Evaluation and Policy Analysis*, 4(4), 461-476. <https://doi.org/10.3102/01623737004004461>
- Smee, S. M., & Blackmore, D. E. (2001). Setting standards for an Objective Structured Clinical Examination: The borderline group method gains ground on Angoff. *Medical Education*, 35(11), 1009-1010. <https://doi.org/10.1111/j.1365-2923.2001.01047.x>
- Kramer, A., Muijtjens, A., Jansen, K., Dusman, H., Tan, L., & van der Vleuten, C. (2003) Comparison of a rational and an empirical standard setting procedure for an OSCE. *Medical Education*, 37(2), 132-139. <https://doi.org/10.1046/j.1365-2923.2003.01429.x>
- Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service.
- Hofstee, W. K. B. (1973). Een alternatief voor normhandhaving bij toetsen. *Nederlands Tijdschrift voor de Psychologie*, 28, 215-227.
- Cohen-Schotanus, J., & van der Vleuten, C. P. M. (2010). A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical Teacher*, 32(2), 154-160. <https://doi.org/10.3109/01421590903196979>
- Wijnen, W. H. F. W. (1971). *Onder of boven de maat*. Amsterdam: Swets & Zeitlinger.

Written By

Gominda Ponnampereuma

MBBS, MMed, PhD

Professor in Medical Education

Faculty of Medicine,

University of Colombo, Sri Lanka

