Health Professions Education

Volume 4 | Issue 4 Article 1

2018-10-18

Standard Setting in Medical Education: Which Cut-Off Cuts it?

Jerome I. Rotgans
Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore,
jerome.rotgans@ntu.edu.sg

Nigel C.K. Tan
National Neuroscience Institute, Singapore

Recommended Citation

Rotgans, Jerome I. and Tan, Nigel C.K. (2018) "Standard Setting in Medical Education: Which Cut-Off Cuts it?," *Health Professions Education*: Vol. 4: Iss. 4, Article 1.
DOI: 10.1016/j.hpe.2018.10.002
Available at: https://hpe.researchcommons.org/journal/vol4/iss4/1

Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education 4 (2018) 233–235

www.elsevier.com/locate/hpe

Standard Setting in Medical Education: Which Cut-Off Cuts it?☆

Standard setting of exams is an important element of medical education. This is understandably so: it is difficult to qualify for medical school in the first place and once there, most assessments are high-stakes and failure rates need to be justified. As a consequence, numerous procedures have been established to determine cut-off values for exams. A cut-off value is the score of an assessment that determines who passes the exam and who does not. The cut-off in turn determines the failure rate, which is the percentage of students below the cut-off value who fail the exam.

There are two main categories of methods to determine the cut-off: criterion-referenced methods and norm-referenced methods.[1] Criterion-referenced methods, which are also referred to as test-centred standard setting, are independent of the test results and are concerned with ascertaining whether the students meet a pre-determined set of requirements.

In practice, this method often entails a panel of experts scrutinising the test items to determine the difficulty level of each item in terms of the hypothetical "borderline" or minimally competent student. A popular method is the Ebel Method in which experts judge each exam item on two dimensions: difficulty (easy, medium, hard) and relevance (essential, important, acceptable, questionable).[2] This procedure is followed by a discussion during which the experts seek to reach agreement among each other on the estimated percentage for each test item. Only after consensus has been reached and the scores have been weighted, the cut-off score can then be determined.

On the other hand, one norm-referenced method establishes its cut-off in the form of a pre-determined percentage of the test questions that need to be answered correctly (e.g., minimum 60% correct). Another basic norm-referenced procedure is to determine the cut-off value by subtracting one standard deviation from the mean. Recently, more sophisticated norm-referenced methods were proposed, such as the Cohen Method and the Modified Cohen Method. In the Cohen Method, the scores of the high-performing students at the 95th percentile are used as the norm reference. In addition, the standard is set to 60% of the top performing students' score and not the cut-off score of correct answers for 60% of the total number of items.[3] A potential shortcoming of the Cohen Method is that the multiplier (.60) and the band of the reference group (95th percentile) were felt to be somewhat arbitrary. The Modified Cohen Method estimates these values more accurately based on historical data from previous exams to come to a more informed decision regarding the final cut-off value (for more details see:)[4]

From the above, one can see that there are many different ways to determine the cut-off values for exams and it appears that there is not one single gold standard. But why is that so? Are the cut-off values generated by these different methods significantly different from each other or do they result in more-or-less the same cut-offs?

If the latter is the case, one does not have to worry that much, as long as one can justify the selection of the method, which will largely be determined by finding the optimum between cost and practicality. Practicality is a significant factor to consider, as faculty time and effort are required to do standard setting for some criterion-referenced methods. However, if these different methods do end up with significantly different results, then it is important to know which one to pick and why.

Since there are not many studies that provide a comparison of a larger variety of cut-off values, generated by different methods, we decided to look at our own data. We extracted the exam data of our first-year medical students for which we have five complete

☆Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region.

https://doi.org/10.1016/j.hpe.2018.10.002

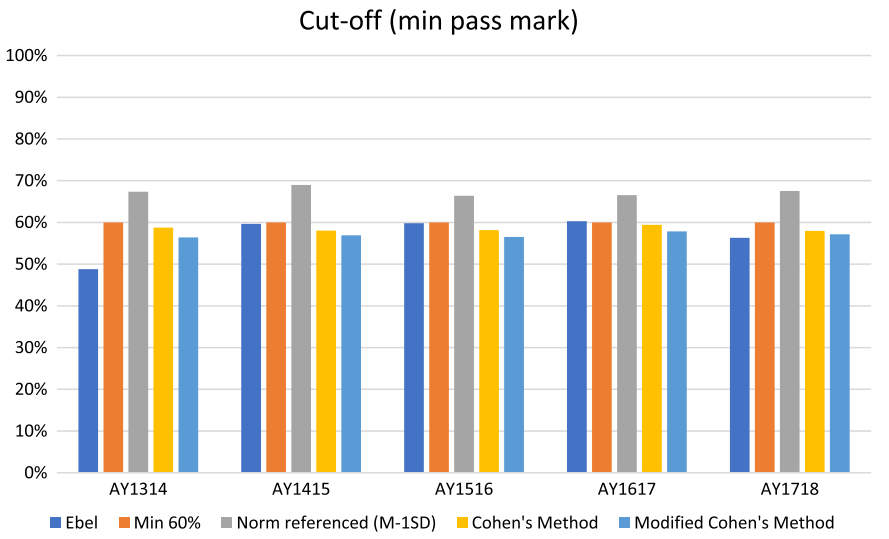

Fig. 1. Cut-off values using (1) Ebel Method, (2) 60% minimal pass, (3) norm referenced, (4) Cohen Method, and (5) Modified Cohen Method, of five cohorts of first-year MBBS students (N = 450).

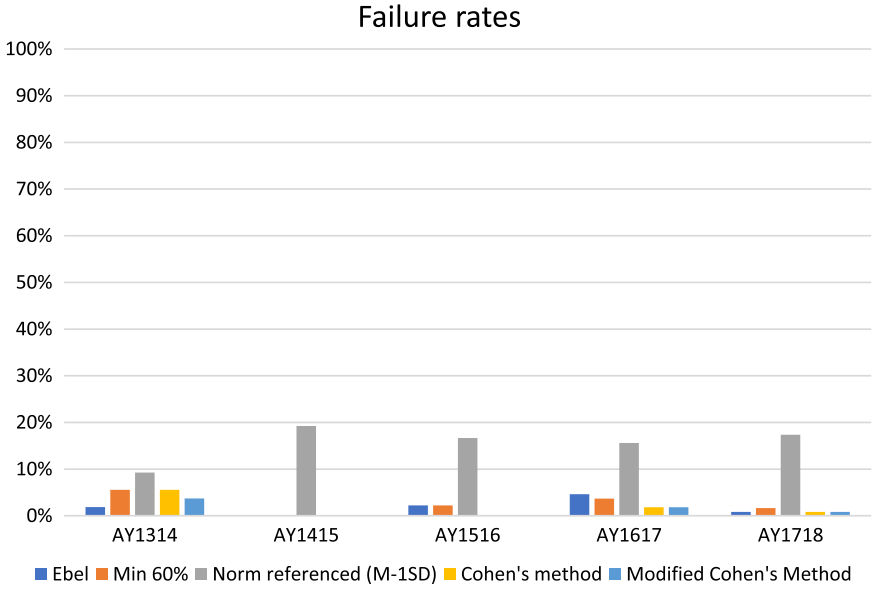

Fig. 2. Failure rates corresponding to (1) Ebel Method, (2) 60% minimal pass, (3) norm referenced, (4) Cohen Method, and (5) Modified Cohen Method, of five cohorts of first-year MBBS students (N = 450).

cohorts and generated the most common cut-off values, as well as the corresponding failure rates for each cohort. The cut-off methods we used were: (1) Ebel Method (note, this is the standard method at the institution the data come from), (2) 60% minimal pass, (3) norm referenced (i.e., Mean − 1SD), (4) Cohen Method, and (5) Modified Cohen Method. The results are depicted in Figs. 1 and 2.

The data revealed that the generated cut-off values are largely similar—except for one, the norm-referenced method (mean − 1SD), which appears to be overly strict. See Cohen-Schotanus and van der Vleuten for an elaboration regarding the issues related to the norm-referenced method, which is not frequently used anymore.[3] The cut-off values for this method are closer to the 70% pass mark, whereas the remaining cut-off values of the other methods are closer to 60%. Unsurprisingly, this more stringent pass mark results in higher failure rates. Taking a closer look at the Ebel Method reveals that it is substantially lower for the first cohort

(AY1213) but does not lead to substantially higher failure rates. The remaining methods are rather stable in terms of cut-off values (the variation across cohorts is less than 1%) and the failure rates are all less than 6%.

Based on our single-centre data from medical undergraduates, it appears that in practice the different methods to determine cut-off values do not result in overly different outcomes (ignoring the norm-referenced method of Mean − 1SD). All methods resulted in more-or-less the same cut-off values around the 60% pass mark and less than 6% failure rates.

Since all methods result in similar outcomes, it is reasonable to conclude that the norm-referenced methods, such as the Cohen Method is perhaps to be preferred over the more time-consuming and thus costlier, criterion-referenced method, such as the Ebel Method. Furthermore, considering that Cohen Method and Modified Cohen Method were also close to the 60% cut-off, one could even go a step further and simply use the a priori determined 60% pass mark as the standard (as it is the case at some Universities). This would save faculty time, since neither time-consuming standard setting procedures are required nor statistical number crunching; just 60% as minimum requirement to pass. Critics, however, would argue that even if we keep to the 60% cut-off, one then has to safeguard that the exams are consistently of the same difficulty level, otherwise the 60% is meaningless. This in turn would require new standards and procedures, which would involve similar variables we use now in the methods for determining the cut-off.

So, are we back where we started? The short answer is, yes. But are there no alternatives? One alternative is: get rid of high-stakes assessments. Note, this does not mean that we should abandon all assessments, but replace high-stakes assessment with frequent low-stakes assessments.

To exemplify this case, the data we reported earlier in this article are taken from a first-year MBBS programme that employs Team-Based Learning (TBL) as its main instructional strategy. In TBL, students have to complete an individual self-test at the beginning of every TBL session to determine whether they have prepared well for the session (known as the individual readiness assurance test). In our school, there are 64 TBL sessions in the first year alone. Each test consists of 25–30 multiple-choice items, which makes together 1,920 items over one year. This constitutes an extensive test with high content coverage and since the test items represent specific learning objectives, the scores should be valid as well.

This might be an alternative to the high-stakes exam that is held at the end of every year. There are some higher-education institutions, outside the field of medicine, that have adapted such an approach and the data show that these frequent, low-stakes assessments produce highly reliable and valid results.[5,6] In addition, students and staff are less stressed and assessment is an integral part of the learning process. We ask: why not adopt it in medical education? Perhaps it is time to explore this further.

References

1. Ben-David MF. AMEE Guide No. 18: standard setting in student assessment. *Med Teach* 2000; **22**(2):120–130.
2. Livingston SA, Zieky MJ. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Services; 1982.
3. Cohen-Schotanus J, van der Vleuten CPM. A standard setting method with the best performing students as point of reference: practical and affordable. *Med Teach* 2010; **32**(2):154–160.
4. Taylor CA. Development of a modified Cohen method of standard setting. *Med Teach* 2011; **33**(12):678–682.
5. Sockalingam N, Rotgans J, Schmidt HG. Student and tutor perceptions on attributes of effective problems in problem-based learning. *High Educ* 2011; **62**(1):1–16.
6. Rotgans JI, O'Grady G, Alwis WAM. Introduction: studies on the learning process in the one-day, one-problem approach to problem-based learning. *Adv Health Sci Educ* 2011; **16**(4): 443–448.

Jerome I. Rotgans*
Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
Institute of Medical Education Research Rotterdam, Erasmus MC, the Netherlands
E-mail address: jerome.rotgans@ntu.edu.sg

Nigel C.K. Tan
National Neuroscience Institute, Singapore
Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

11 October 2018; accepted 14 October 2018

*Corresponding author at: Nanyang Technological University, Lee Kong Chian School of Medicine, 11 Mandalay Road, 308232, Singapore.