

Standard Setting for Assessments in Anatomy using the Angoff Method in a Medical School in Sri Lanka

C S Paththinige¹, S M A Jayawardana¹

¹Department of Anatomy, Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka

Abstract

Introduction: Anatomy is a fundamental subject in medical education. A reliable and valid assessment tool is crucial for evaluating students' understanding. The Angoff method, an absolute method of standard-setting has been widely used in various educational contexts. This study aimed to develop and implement a standard-setting process for an assessment in Anatomy using the Angoff method in a Sri Lankan medical school.

Methods: Five Anatomy faculty members including two senior lecturers and three temporary assistant lecturers who were recent graduates participated in the standard-setting process. They were presented with each item on the anatomy assessment and asked to estimate the percentage of minimally competent students who would answer the item correctly. Their judgments were then used to determine the cut score for the assessment.

Results: The Angoff method yielded a cut score of 56.06 with a cutoff value of 8.41 and a standard error of judgment of 0.18 (1.2%). There was a high interrater reliability (0.89).

Conclusion: This study demonstrated the feasibility and effectiveness of using the Angoff method for standard setting in an anatomy assessment in a Sri Lankan medical school. The set cut score provides a benchmark for evaluating students' performance and ensuring that they possess the necessary anatomical knowledge to progress in their medical education.

DOI: <https://doi.org/10.4038/slaj.v8i2.254>

Sri Lanka Anatomy Journal 2024; 8(II): 27–35

Corresponding author

C S Paththinige

E-mail: pcsampath@med.rjt.ac.lk



<https://orcid.org/0000-0001-9214-7808>

SLAJ, 8(II) 2024

CC BY 4.0



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 license (unless stated otherwise) which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. Copyright is retained by the author(s).

Introduction

In medical education, assessments are used to determine if the trainees/students are meeting the required standards of medical competence for clinical practice and, therefore play a crucial role in ensuring the quality of medical education. Assessments allow educators to evaluate the competencies required of students to progress to the next stage in their educational program. Moreover, assessments provide essential feedback to educators and students to identify the learning gaps, engage in self-reflection, adjust their teaching or learning strategies, and provide motivation for future learning (1,2). Standards-setting, a critical component of assessment, involves determining the cutoff point that distinguishes competent from non-competent examinees. This process ensures fairness, transparency, and serves as a quality control mechanism (3).

The standards can be either ‘norm-referenced’ (based on the performance of a large representative reference group) or ‘criterion-referenced’ (based on the predetermined level of competence). The standards can also be categorized into ‘relative standards’, which focus on the performance of the examinees as a group, and ‘absolute standards’, which focus on the achievement of a predetermined level of competence by the examinees. Further, the standard setting methods are classified into ‘test-centered’ methods and ‘examinee-centered’ methods, which focus on the individual test items, and the performance of examinees respectively (3,4). Based on these classifications, four broader groups emerge: Relative methods, Absolute methods based on

test questions, Absolute methods based on examinees, and Compromise methods (4).

- **Relative methods:** These methods select a certain number of examinees, often used in admission or placement tests. An example is the fixed percentage method, where a panel of experts sets a fixed percentage of examinees as passing, regardless of their actual performance (4).
- **Absolute methods based on judgments about test questions:** In these methods the experts judge the difficulty of individual test items and set the standard accordingly. The Angoff Method is a common example, where experts estimate the percentage of minimally competent (“borderline”) examinees who would answer each item correctly (3,4).
- **Absolute methods based on judgments about individual examinees:** Experts judge the performance of individual examinees to set the standard. For example, in the contrasting group method, experts evaluate performance of two groups (high and low performers) to determine the cutoff point (3,4).
- **Compromise methods:** These methods combine elements of relative and absolute methods. The Hofstee Method sets a standard based on a fixed percentage (failure rates) and expert judgment (on difficulty of items). The standard is set where the predicted pass/fail rates match the predetermined fixed percentage (3,4).

Educators should decide on the suitable standard-setting method considering the purpose of the examination, available evidence, and the contextual factors that

determine the feasibility of employing each method in their educational setting (4).

A well-defined standard in assessment not only provides a reliable indicator of the student's level of competency, but also provides feedback on the assessments and the educational programs in general and supports informed decision-making for the improvement of the educational programs (3). Therefore, this activity was conducted from a test development perspective aiming for the improvement of the assessment systems in Anatomy in Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka (FMAS, RUSL).

Methods

The six-step approach for 'setting standards on written examinations' described by Norcini (4) was used as the framework for this task. These six steps are,

1. Deciding on the type of standards
2. Deciding on the method for setting standards
3. Selecting the judges
4. Holding the standard setting meeting
5. Calculating the standard/ cutoff point
6. Establishing what to do afterward

The activities done in setting standards for the Assessment in Anatomy, using scenario-based single-best answer multiple choice questions (MCQs) under these six steps are described below.

1. Deciding on the type of standards:

The assessment of Anatomy with scenario-based single-best answer MCQs aims for an assessment of higher-order thinking skills, such as application and analysis, rather than factual recall. 'Absolute standard' that is based on a predetermined level of competency is selected as the appropriate type of standard in this task because absolute standards ensure a consistent level of competence regardless of the performance of other examinees, therefore provide a clearer indication of whether a candidate has truly acquired the required level of competency (3,4).

2. Deciding on the method for setting standards

Among four groups of standards-setting methods described by Livingston and Zeiky, as cited by Norcini, (i.e. Relative methods, Absolute methods based on judgment about test questions, Absolute methods based on judgment about examinees, and Compromise method) (4), absolute methods based on judgment about test questions were used as the most appropriate method in this activity, because one of the main aims of this standard-setting task was to improve the test items. The most commonly used standard-setting methods in this category are Angoff's method and Ebel's method. The standard setting in Angoff's method is based on the estimates of the percentage of borderline candidates who would answer each item correctly. In Ebel's method, items are categorized into groups based on their difficulty level (easy, medium, hard) and relevance (essential, important, acceptable), and expert judgment on the

percentage of correct answers for each category by a borderline candidate is obtained. Ebel's method relies more on the judgment of experts (for categorization of items), which is a limiting factor in human-resource limited settings like ours. Angoff's method is considered more tiresome for the standard setters because it involves the evaluation of each item in isolation (3). However, in this activity the assessment consisted of only 15 items, and the aim was to obtain judgment on each item for the purpose of improvement, therefore Angoff's method is selected as the most suitable method for this task.

3. *Selecting the judges*

Considering the low stake setting in our assessment, five judges: two senior lecturers in Anatomy and three temporary assistant lecturers in the Department of Anatomy, FMAS, RUSL (recent medical graduates) were selected.

4. *Holding the standard setting meeting*

A meeting was held with the participation of all five judges discussing the purpose and the methods in detail. The nature of the expected competence (mainly application of knowledge) and the characteristics of a 'borderline examinee' were discussed and agreed upon.

5. *Calculating the standard/ cutoff point*

The first two items were rated independently by the five judges and a discussion was held to ensure that each rater had comprehended the method. The next 13 items were rated individually by the five judges. The estimates

of the five judges on all five items were entered into the 'Angoff Analysis Tool' which is a MS Excel spreadsheet that is specifically designed to analyze the results of Angoff's studies obtained from the Assessment Systems Corporation (ACS) (5). A second round of ratings was not conducted because none of the item ratings had a standard deviation (SD) over 10.

6. *Establishing what to do afterward*

The results were presented to the judges, and they agreed on the credibility of the results.

The methodology of this standard setting exercise is illustrated in the Figure 1.

Results

Ratings of the 15 items of the assessment by five judges in Angoff's analysis is presented in Table 1.

The output of the 'Angoff Analysis Tool' with the results of the Angoff analysis is presented in Table 2. The Angoff analysis provided a cutoff value of 8.41 (56.06%), with an Interrater reliability-based standard error/ Standard error of judgment of 0.18 (1.2%). There was a high IRR (0.89)

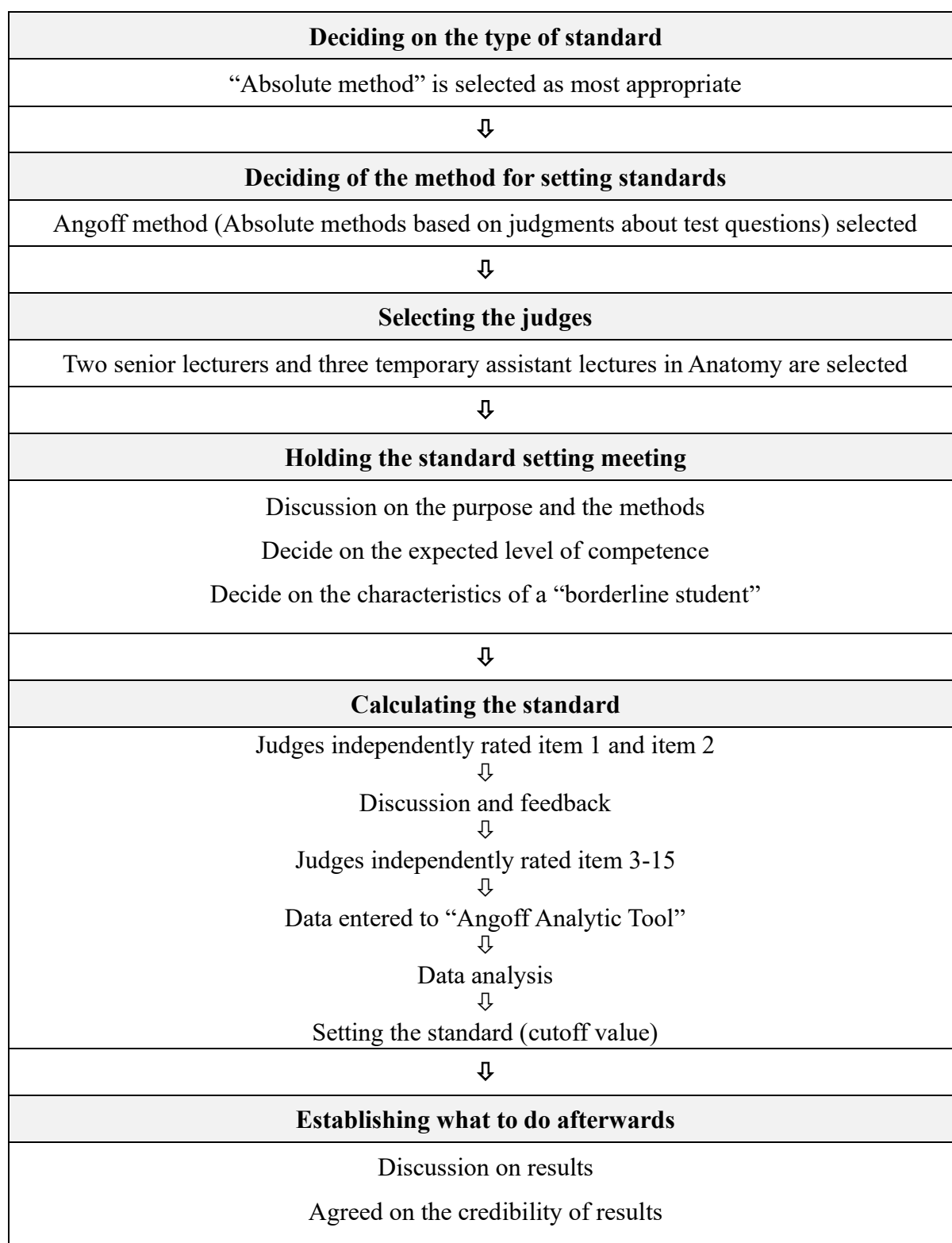


Figure 1: Outline of the methodology of the standard setting exercise

Based on the six-step approach for ‘setting standards on written examinations’ by Norcini (4)

Table 1. Ratings of the test items by five judges in Angoff's analysis

Item	Rating by the judges					Average	SD	Min.	Max.
	1	2	3	4	5				
1	70	50	60	70	50	60.00	8.94	50	70
2	50	50	50	60	50	52.00	4.00	50	60
3	80	70	60	70	60	68.00	7.48	60	80
4	60	50	50	50	50	52.00	4.00	50	60
5	50	60	60	55	50	55.00	4.47	50	60
6	40	60	60	50	40	50.00	8.94	40	60
7	50	50	50	50	50	50.00	0.00	50	50
8	50	60	50	55	50	53.00	4.00	50	60
9	60	70	60	70	60	64.00	4.90	60	70
10	40	40	50	40	40	42.00	4.00	40	50
11	70	70	70	65	60	67.00	4.00	60	70
12	50	50	50	55	40	49.00	4.90	40	55
13	80	80	60	75	60	71.00	9.17	60	80
14	70	70	60	65	50	63.00	7.48	50	70
15	40	50	50	45	40	45.00	4.47	40	50
Average	57.33	58.67	56.00	58.33	50.00	56.07			
SD	13.40	10.87	6.11	10.11	7.30				

Max.- Maximum rating, Min.- Minimum rating; SD- Standard deviation

Table 2. Results of Angoff analysis

Statistics	Raw scale	% scale
Average:	8.41	56.07
SD:	0.53	3.55
Minimum:	7.5	50
Maximum:	8.8	58.67
Sample SE:	0.24	1.59
IRR-based SE:	0.18	1.17

IRR	0.89
-----	------

Cut - SEJ	Cut	Cut + SEJ
8.23	8.41	8.59
54.87%	56.07%	57.27%

IRR- Interrater reliability; SD- Standard deviation; SE- Standard error
SEJ- Standard error of judgment

Discussion

This study successfully implemented an Angoff-based standard-setting process for an Anatomy assessment at the Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka. Despite the limited resources available, the process provided valuable insights into the assessment's difficulty and helped to ensure a fair and equitable evaluation of student performance. The findings revealed that the selected items were of appropriate difficulty for the target population. The panel's judgments were consistent and reliable, demonstrating the effectiveness of the Angoff method in this context.

The Angoff and modified Angoff methods are commonly employed absolute standard-setting procedures for MCQs. While resource-intensive, these methods provide a defensible, credible, and reliable framework for establishing performance benchmarks. The incorporation of expert judgment, informed by their experience, and the flexibility offered by various modifications enhance the validity and utility of these approaches in educational and assessment contexts (6).

Involvement of adequate number of subject matter experts in the panel of judges is the main challenge experienced in this activity, which is a common limitation experienced by educators in resource limited settings. Subject matter experts are described as individuals

who possess a deep understanding of the subject matter being assessed, which often requires advanced education/training and/or significant professional experience in the field. A large-scale study with 4900 simulated panels recommended the involvement of judges with mixed expertise to yield optimal results in standard setting using Angoff or Modified Angoff methods (7). Securing the participation of experts can be difficult due to their busy schedules and commitments because the Angoff process can be time-consuming, requiring multiple rounds of judgment and discussion. Expert panelists may be influenced by their overall impression of an item or an examinee, rather than focusing solely on the specific item's difficulty. This can lead to biased judgments. Therefore, medical educators have tried alternative methods in Angoff procedure such as involving the recent graduates or postgraduate students in the absence of experts with convincing outcomes (8,9).

In a personal communication with the first author, Professor John Norcini endorsed the use of recent graduates as raters in Angoff method of standard setting for basic sciences examinations. The recent graduates as the 'real consumers' of the curriculum and with their familiarity with the characteristics and abilities of the target student population, such as the students' educational background, cultural context, and potential challenges were described as the main strength of this group, and they are less likely to be influenced by preconceived notions (8). However, they lack experience in standard-setting procedures, that required additional training and guidance to ensure accurate and consistent judgments.

Having multiple panelists can help to reduce the impact of individual biases and increase the reliability of the standard-setting process (4). However, there are no specific recommendations on the number of experts in the panel of judges in Angoff or Modified Angoff methods. Various studies have involved less than five to more than 25 subject matter experts in their standard setting activities with variable results. Ten to fifteen judges have been commonly described as the optimal range, still a fewer number would be sufficient for low stake examinations (10). Irrespective of the number of experts and the level of expertise of the judges, it is essential to employ statistical methods to analyze and reconcile differences in panelists' judgments (11,12). The credibility of the results can be further enhanced by making the judges aware of the results of the pilot test (e.g. overall mean and standard deviation) and individual item analysis (e.g. difficulty index), and holding a second (or third) round of judging if significant differences in opinion are observed in the initial round (4).

In conclusion, the study provided valuable feedback for improving future assessments. The identified areas for improvement include, providing additional training to judges to enhance their understanding of the Angoff method and improve the consistency of their judgments, involving experts from other institutions and allocating more resources to the standard-setting process to ensure a more robust and comprehensive evaluation. Overall, this study demonstrates the feasibility of conducting a successful standard-setting process in resource-limited settings. The findings have implications for improving the

quality and fairness of assessments in higher education institutions, particularly in developing countries.

References

1. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007; 356(4): 387-396.
2. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, International CBME Collaborators. The role of assessment in competency-based medical education. *Med Teach*. 2010; 32(8): 676-682.
3. Ben-David MF. AMEE Guide No. 18: Standard setting in student assessment. *Med Teach*. 2000; 22(2): 120-130.
4. Norcini JJ. Setting standards on educational tests. *Med Educ*. 2003; 37(5): 464-469.
5. Assessment Systems Corporation. (2023) The Angoff Analysis Tool. Available at: <https://assess.com/angoff-analysis-tool/> (Accessed 02 January 2024)
6. Saaiq M. Standard Setting Methods for the Assessment of Knowledge and Skills in Medical Education. *J Health Prof Edu Innov*. 2024; 1(2): 14-20.
7. Shulruf B, Wilkinson T, Weller J, Jones P, Poole P. Insights into the Angoff method: results from a simulation study. *BMC Med Educ*. 2016; 16: 134.
8. Verhoeven, der Steeg V, Scherpbier, Muijtens, Verwijnen, Der Vleuten V. Reliability and credibility of an Angoff standard setting procedure in progress testing using recent graduates as judges. *Med Educ*. 1999; 33(11): 832-837.
9. Mubuuqe A, Mwesigwa C, Kiguli S. Implementing the Angoff method of standard setting using postgraduate students: Practical and affordable in resource-limited settings. *Afr J Health Prof Educ*. 2017; 9(4): 171-175.
10. Hurtz GM, Hertz NR. How many raters should be used for establishing cutoff scores with the Angoff method? A generalizability theory study. *Educ Psychol Meas*. 1999; 59(6): 885-897.
11. Maurer TJ, Alexander RA, Callahan CM, Bailey JJ, Dambrot FH. Methodological and psychometric issues in setting cutoff scores using the Angoff method. *Pers Psychol*. 1991; 44(2): 235-262.
12. Tavakol M, O'Brien D, Stewart C. Determining intra-standard-setter inconsistency in the Angoff method using the three-parameter item response theory. *Int J Med Educ*. 2023; 14: 123-130.