A New Method for Setting Standard in Medical Education, Applying Previous Year Cumulative GPA

Myo Nyein Aung, Vorapol Jaroonvanichkul, Jitladda Deerojanawong, Juraiporn Somboonwong, Ishtiaq Ahmad, and Pongsak Wannakrairot

Abstract—This study tested a new method of standard setting which does not need judges’ decision. Objectives: (1) to compare the minimum-passing-level (MPL) of the new standard setting approach with that of a traditional method (Angoff’s Method); and (2) to find the best borderline group by analyzing five ranges of the students’ previous year’s Cumulative Grade Point Average (cGPA). Method: A range of previous year GPA around 2 was used to probe borderline group, the mean of which was applied as cut-off to decide pass and fail. Receiver operating characteristic (ROC) curve and area under the curve (AUC) analyses were applied to investigate the agreement between the pass/fail cut-scores of the new approach and Angoff’s Method. Result: The pass/fail cut-score of new method highly agreed MPL of Angoff’s Method: with AUC ranging from 0.91 to 1.0 in all analyses. Conclusion: This new approach would be an alternative to Angoff’s Method when subject-matter experts are not available or in order to avoid the making of subjective decisions by these experts.

Index Terms—MPL, Pass and Fail, Medical Education, Measurement, Standard Setting.

I. BACKGROUND

After all medical school examinations, medical teachers decide whether students pass or fail. However, limitation of resources can avert the proper standard setting approach for such decision, especially within resource limited settings. A method which is practicable without the burden of technical and human resource needs as well as being comparably standardized to existing well-known methods is still a need for medical education measurement. This study introduces an innovative student-centered standard setting (SCSS) approach which rates favorably in comparison to the widely used Angoff’s Method.

Test-centered standard setting approaches such as Angoff’s Method (since 1971) Ebel’s Method (since 1972) and Nedelsky’s Method (since 1954) have been the most commonly studied methods in the field of medical education [1]. Yet there exists no perfect setting standard with the different methods producing varying passing scores and there is no gold standard in setting standard [2]. Every method represents a methodological attempt to determine an unbiased differentiation of the minimally competent level in an assessment of student competency, performance or knowledge. Experts indicate that the key to defensible standard lies in the creditable decisions of test graders along with the use of systematic approaches by them to reach their decisions [3].

Another group of standard setting methods known as examinee-centered approaches such as Cohen’s method, Hofstee’s Method, the Contrasting Groups Method, the Borderline Group Method, and the Generalized Examinee-Centered Method have similarly been well studied within medical education [4]. These methods additionally may require panelists to review the examinee’s work, as in the Contrasting Group Method, to categorize the examinees, as in the Borderline Group Method, and to evaluate using a rating scale, as in the Generalized Examinee-Centered Method. A student-centered approach which avoids the need for judges, as required with Angoff’s method, and escaping a fixed percentage of failed students, as is the case in mean and standard deviation (SD) methods would be a practical option for many medical schools.

A borderline group is a group comprising students who have a borderline level of competence in any given test [2]. Yet determining who constitutes this borderline group is a practical challenge [5]. In this study, we tested a new approach for defining the borderline group of students using the students’ cumulative grade point average (cGPA) of preceding year. In this new approach of standard setting, students’ cGPA earned in the previous academic year was used a probe to identify a sample of borderline group.

Applying students’ cGPA’s to determine an exam’s minimum passing level (MPL)

A student’s cGPA is reflective of his or her abilities through the preceding academic years [6]. Score grades reflect a student’s abilities via assessment of his or her performance. These scores indicate whether students have acquired competencies through the interaction of their abilities with the difficulty of the assessment questions.

A student-centered approach to standard-setting will endeavor to form a borderline group based on the abilities of the whole class of students. The cGPA of the class in previous year can be used to classify the students. In this...
approach, the borderline group is formed by students appearing within the grey zone as defined by a range of the cGPA earned in the previous year, such as GPA 1.5 to 2.5, 1.6 to 2.6, 1.7 to 2.7, 1.3 to 2.3, or 1.4 to 2.4. The minimum passing level (MPL) for each examination is set as the mean score of the students in the borderline group. In this research, the MPL determined by five borderline groups were compared with the MPL’s as determined by a panel of experts according to Angoff’s method with the aims of (1) to compare MPL of a new standard setting method with MPL of traditional Angoff’s method in 4 examination data sets. (2) to identify the best borderline group applying five ranges of previous years cumulative GPA (cGPA).

II. Method

The Faculty of Medicine, Chulalongkorn University, registry department allowed for the use of students’ record data. The method was tested with 4 sets of data. The data sets were those of pathology summative examinations in the years 2009, 2010 and 2011. The items of the question set underwent difficulty tests and item analysis before administering each examination.

A. Concept of the new method

The distribution of raw scores in an assessment test (examination) reveals the interaction between the competencies of the students and the assessment test. Therefore, the score of those borderline group students may serve as a guide to decide the minimal passing level.

The ability of the students was taken from the students’ cGPA of the previous year.

B. Forming Borderline groups

Based on preceding year cGPA, students who got cGPA level of 2 were considered as the with minimally sufficient competencies. A cGPA of less than 2 was regarded as referring to students who do not have sufficient performance. Therefore, a range of cGPA around 2 was used to form the borderline group of students.

The five borderline groups were organized as follows: borderline group 1 included students whose cGPA was 1.5 to 2.5; the borderline group 2 cGPA range was 1.4 to 2.4; the borderline group 3 cGPA range was 1.3 to 2.3; the borderline group 4 cGPA was 1.6 to 2.6; and the borderline group 5 cGPA range was 1.7 to 2.7. (Table Ia to Id).

C. Defining the cut-off points MPL

Given that the assessment tools for the target assessment will have been standardized, the mean score of the students in the borderline group in each test can be applied as MPL. Therefore, the raw scores of the borderline group, defined by the previous year cGPA, were averaged to get the mean value. The mean of the test score or examination mark in each BG was taken as the MPL so determining the pass and fail cut-off point.

To fulfil the objective 1, MPL of the new method was compared to MPL of Angoff’s method in term sensitivity, specificity and Receiver Operating Characteristics (ROC) curve analysis. To fulfill the objective 2, different ranges of borderline groups were tested, including students in the cGPA ranges of 1.5 to 2.5, 1.6 to 2.6, 1.7 to 2.7, 1.4 to 2.4, and 1.3 to 2.3 comparing Area under ROC curves (AUC).

D. Data set

Four data sets of raw scores from pathology summative assessment tests were applied in this study to compare the
MPL’s of the new method with those from Angoff’s Method. The examination sets related to the principles of pathology (2011 and 2012) and the endocrine and genitourinary system (2010 and 2011). These examinations applied multiple choice questions (MCQ).

E. Analysis

The analysis comprises ROC analysis in four data sets. The graphical plot ROC curve will illustrate the performance of standard to discriminate pass and fail, when the cut-off point is varied. X axis is showing true positive rate in term of sensitivity and Y axis is showing false positive rate in term of (1-specificity).

In each analysis, the MPL of five borderline groups were compared to the pass and fail cut off points generated by a panel of experts in Angoff’s approach. Comparing performance of new method MPL against that of Angoff’s, sensitivity, specificity and ROC curves were produced. Meanwhile, ROC curves of five BGs were compared in term of AUC. When AUC approached to 1 upon being compared to the pass and fail cut off point of Angoff’s method, it meant that the new method was a best BG.

The MPL applying Angoff’s Method was 30.48. According to this MPL, 297 (94.59%) students passed the exam. The MPLs of Angoff’s Method and the new method were then compared in terms of sensitivity, specificity and ROCs (Table Ia). The ROC and AUC results for five borderline groups are shown in Table Ia and Fig. 2.

III. RESULTS

Results in each data set are presented for four datasets. Performance of MPL decided by new standard setting method produced almost identical results as Angoff’s. (Table Ia to Table Id and Fig. 2 to 4). Moreover, comparing five MPLs formed by five BGs using cGPA around 2, the best BG was probed by cGPA 1.3 to 2.3 and it yielded MPL, most identical to that of Angoff’s Results for two objectives in each examination data set are as followed.

A. Result of data set 1

Data set 1 represents examination scores from 2012 of 314 students in an assessment of the Principles of Pathology, this was a multiple choice type of assessment comprising 59 questions. The maximum obtainable score was 59. The MPL applying Angoff’s Method was 30.48. According to this MPL, 297 (94.59%) students passed the exam.

The borderline group, using the students’ previous year’s CGPA range 1.3 to 2.3, comprised 28 students. The MPL of the new method (33.04) was applied to the students in the borderline group. Applying this MPL, 275 (87.58%) students passed the exam.

TABLE I. 

<table>
<thead>
<tr>
<th></th>
<th>Borderline group</th>
<th>AUC</th>
<th>95% CI</th>
<th>sensitivity</th>
<th>95% CI</th>
<th>specificity</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GPA 1.5 to 2.5</td>
<td>0.9512</td>
<td>0.93-0.97</td>
<td>90.2</td>
<td>86.3-93.4</td>
<td>100</td>
<td>80-100</td>
</tr>
<tr>
<td>2</td>
<td>GPA 1.4 to 2.4</td>
<td>0.9512</td>
<td>0.93-0.97</td>
<td>90.2</td>
<td>86.3-93.4</td>
<td>100</td>
<td>80-100</td>
</tr>
<tr>
<td>3</td>
<td>GPA 1.3 to 2.3</td>
<td>0.9630</td>
<td>0.95-0.98</td>
<td>92.6</td>
<td>89-95.3</td>
<td>100</td>
<td>80-100</td>
</tr>
<tr>
<td>4</td>
<td>GPA 1.6 to 2.6</td>
<td>0.9512</td>
<td>0.93-0.97</td>
<td>90.2</td>
<td>86.3-93.4</td>
<td>100</td>
<td>80-100</td>
</tr>
<tr>
<td>5</td>
<td>GPA 1.7 to 2.7</td>
<td>0.9125</td>
<td>0.89-0.93</td>
<td>82.5</td>
<td>77.7-86.6</td>
<td>100</td>
<td>80-100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Borderline group</th>
<th>AUC</th>
<th>95% CI</th>
<th>sensitivity</th>
<th>95% CI</th>
<th>specificity</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GPA 1.5 to 2.5</td>
<td>0.9683</td>
<td>0.95-0.98</td>
<td>93.7</td>
<td>90-96.3</td>
<td>100</td>
<td>79-100</td>
</tr>
<tr>
<td>2</td>
<td>GPA 1.4 to 2.4</td>
<td>0.9776</td>
<td>0.97-0.99</td>
<td>95.5</td>
<td>92.9-97.7</td>
<td>100</td>
<td>79-100</td>
</tr>
<tr>
<td>3</td>
<td>GPA 1.3 to 2.3</td>
<td>0.9832</td>
<td>0.97-0.99</td>
<td>96.6</td>
<td>93.7-98.5</td>
<td>100</td>
<td>79-100</td>
</tr>
<tr>
<td>4</td>
<td>GPA 1.6 to 2.6</td>
<td>0.9496</td>
<td>0.93-0.97</td>
<td>89.9</td>
<td>85-93.3</td>
<td>100</td>
<td>79-100</td>
</tr>
<tr>
<td>5</td>
<td>GPA 1.7 to 2.7</td>
<td>0.9366</td>
<td>0.92-0.96</td>
<td>87.3</td>
<td>82.9-91.1</td>
<td>100</td>
<td>79-100</td>
</tr>
</tbody>
</table>

AUC = Area under the ROC curve

DOI: http://dx.doi.org/10.24018/ejmed.2019.1.5.114
Five borderline groups using different ranges of GPA around GPA2 were tested. The AUC for all borderline groups was more than 0.9 (Fig. 2). The MPL of the borderline group which comprised a CGPA range of 1.3 to 2.3 achieved the largest AUC (0.963). (Fig. 2)

The specificity of MPL determined by the new cGPA-based approach is 100% when compared to the pass-fail decisions of Angoff’s Method. There was a slight variation in sensitivity among the MPLs determined by the different CGPA ranges. The accuracy of MPLs is best seen when the areas under each ROC curve are compared to each other, with the largest AUC representing the best performance. MPLs obtained from the borderline group using a CGPA range of 1.3 to 2.3 yielded the largest AUC (0.963).

B. Result of data set 2

Data set 2 represents examination scores from 2010 of 284 students in an assessment of the endocrine and genitourinary system pathology. This was a multiple-choice type of assessment comprising 60 questions. The maximum obtainable score was 100. The MPL applying Angoff’s Method was 52.05. According to this MPL, 268 (94.37%) students passed the exam.

The borderline group, using the students’ previous year’s CGPA range 1.3 to 2.3, comprised 37 students. The MPL of the new method (56.76) was applied to the students in the borderline group. Applying this MPL, 259 (91.20%) students passed the exam. The MPLs of Angoff’s Method and the new method were then compared in terms of sensitivity, specificity and ROC curves (Table Ia). The ROC and AUC results for five borderline groups are shown in Table Ib and Fig. 3.

Five borderline groups using different ranges of GPA around GPA2 were tested. The AUC for all borderline groups was more than 0.9 (Fig. 3). The MPL of the borderline group which comprised a CGPA range of 1.3 to 2.3 achieved the largest AUC (0.983). (Fig. 3)

The specificity of MPL determined by the new CGPA-based approach is 100% when compared to the pass-fail decisions of Angoff’s Method. There was a slight variation in sensitivity among the MPLs determined by the different CGPA ranges. The accuracy of MPLs is best seen when the areas under each ROC curve are compared to each other, with the largest AUC representing the best performance. MPLs obtained from the borderline group using a CGPA range of 1.3 to 2.3 yielded the largest AUC (0.9832).

C. Result of data set 3

Data set 3 represents examination scores from 2011 of 292 students in an assessment of the principle of pathology. This was a multiple-choice type of assessment comprising 60 questions. The maximum obtainable score was 60. The MPL applying Angoff’s Method was 30.86. According to this MPL, 253 (86.64%) students passed the exam.

The borderline group, using the students’ previous year’s CGPA range 1.3 to 2.3, comprised 41 students. The MPL of the new method (30.87) was applied to the students in the borderline group. Applying this MPL, 253 (86.64%) students passed the exam. The MPLs of Angoff’s Method and the new method were then compared in terms of sensitivity, specificity and ROCs (Table Ia). The ROC and AUC results for five borderline groups are shown in Table Ic.

Five borderline groups using different ranges of GPA around GPA2 were tested. The AUC for all borderline groups was more than 0.9. The MPL of the borderline group which comprised a CGPA range of 1.3 to 2.3 achieved the largest AUC (1.00).
The specificity of MPL determined by the new CGPA-based approach is 100% when compared to the pass-fail decisions of Angoff’s Method. There was a slight variation in sensitivity among the MPLs determined by the different CGPA ranges. The accuracy of MPLs is best seen when the areas under each ROC are compared to each other, with the largest AUC representing the best performance. MPLs obtained from the borderline group using a CGPA range of 1.3 to 2.3 yielded the largest AUC (0.971).

Analyses of the four sets of examination data showed consistent results despite the nature of the subjects within the four examination datasets and the years of the students being different.

IV. DISCUSSION

The concept of competence is increasingly highlighted in twenty-first century education [7]. Therefore, setting the assessment standard should seek to judge the professional value in terms of a student’s competence for carrying out a particular educational outcome. To test competence, such as in final or exit examinations, and tests for certification and licensure, absolute standards are more appropriate in determining whether examinees have sufficient knowledge or are competent enough to carry out such an educational outcome [8].

Several different methods for setting standards are described in medical education literature. They all require a group of panelists [9] however, which cannot be practically feasible in many settings of the world. Moreover, today medical teachers require a faster, more accurate, and less human-resource-intensive approach for setting standards. In order to ensure a global standard for medical graduates, a practicable student-centered approach is necessary. Such a student-centered approach for deciding the minimum passing level for an examination, however, has yet to be designed (8). In this study, a new approach towards standard setting was tested against Angoff’s method, the widely accepted and applied method within medical schools globally [10]-[11]. Statistical agreements were established by ROC curves. The study yielded promising results.

ROC analysis is a very good tool to measure the agreement between the performance of new method cut-off points and Angoff’s method cut-off point [3]. ROC analysis [12] has been the best fit statistical analysis in epidemiological diagnostic research [13] and is becoming increasingly popular among other disciplines such as psychometric and education measurement. The strength of ROC analysis is being able to visualize the agreement of new tests with a reference test within ROC curves and the areas under ROC curves (AUC) [14]-[15]. The application of ROC analysis was recommended in recent medical education literature by Tavakol and Dennick to see how effectively a cut score can discriminate between passing and failing students [3]. AUC represents how well a test can differentiate people with disease and from people without disease. Likewise, AUC applied in standard setting will inform how accurately it can differentiate a student with required level of competence from those who are below the level of MPL. The larger the AUC, the better the differentiation among passing and failing students [3].

Current study finding revealed the performance of a new method in a high agreement to Angoff’s with large AUC. The borderline group sampled by cGPA 1.3 to 2.3 was identified as the best group to apply as the range of cGPA to form borderline group. The results were found the same upon repeating the test and analyses in four examination data sets.

Moreover, choosing students who will form the...
borderline group in a class is challenge for standard setting. The concept of selecting borderline students in this new approach may mimic idea of sampling in epidemiological survey. Given the whole class is a population, then the borderline group is a sample which will represent students who may have obtained minimal level of score in an assessment test. Results in this study showed that applying cGPA earned in the previous year can be a good probe to get a sample of borderline study from the whole class. Using a range of cGPA 1.3 to 2.3 as a probe, teachers can easily identify a sample of students with borderline level of required competences to pass an assessment test. Efficacy of this new approach was well proven by testing in four sets of data. (Table 1a to 4, Fig. 1 to 4) The findings from the analyses in this study consistently indicated that the two approaches yielded almost identical cut-off points when determining pass and fail levels (Table 1a to Id).

Another strength of this the new method was that it overcame the human resource intensiveness of Angoff’s Method whereas it could produce a similar MPL. It is simple to apply as it does not require complex statistical skill to get MPL.

Obviously, this new approach does not require a panel of experts to decide the minimal passing level. All that is required is a simple statistical analysis based on the cGPA of the students from the previous year. This method, therefore, would be useful in places where the number of experts required to apply Angoff’s Method or other standard setting methods is not available [16]. It would be of particular use in medical schools with large class sizes to produce assessment results in a standardized and speedy way.

A number of MPL ranges were tested in order to obtain a best fit curve that agreed with the MPL cutoff point of Angoff’s Method (Fig. 1-4). A cGPA level of 2 was considered as the indicator of students with minimally sufficient competencies from the previous year. Consequently, this score may indicate the minimal passing level given that the assessment tools for the target assessment will have been standardized. Therefore, it was considered logical to form the borderline group based on a cGPA of 2. The range included a cGPA of 2.

Test cut scores are usually defined in norm-referenced or criterion referenced approaches. In line with a student-centered standard setting approach the ability of the students and the competency in the current assessment examination were considered when forming the borderline group (BG). It is to be cautious that we did not try to invent a new theory. We implied an approach which would alternatively be applicable as Angoff’s whenever there is resource limitation.

In this approach, the mean score of the borderline group is required to set the MPL. However, medical schools with large class sizes may use the median instead of the mean [17]. Moreover, the examination score sets used in this data analysis were solely those of multiple choice questions. Future research to test this new method of standard setting in clinical examinations may be required.

In this study scores from multiple pathological examination result from one faculty was repeated trialed to test the new method. Future researches may validate this new approach with other score from other departments and in various medical schools.

V. CONCLUSION

The GPA of a student is the key indicator reflecting academic performance. Nowadays most universities use the GPA system. Therefore, defining borderline groups applying the cGPA represents a practical, student-centered approach to assessment standard setting. We believe that this students-centered standard setting (SCSS) approach may enable medical school faculties to take evidence-based pass and fail decisions for their students with a reduced time, financial and human resource burdens.

VI. ETHICS AND CONSENT TO PARTICIPATE

Three was no collection of data from human subjects. The examination score was obtained with the official permission from registry department for ethics approval. All authors agreed to publish the finding and the article.

VII. CONFLICT OF INTEREST

No competing interest in this research.

ACKNOWLEDGMENT

Pongsak Wannakriarot and Myo Nyein Aung designed the study. Myo Nyein Aung developed the conceptual framework, carried out the analysis and wrote the manuscript, taking into account all co-authors’ contributions of advice and comments. Jitladda Deerojanawong, Juraiporn Somboonwong and Vorapol Jaroonvanichkul performed the data collection, obtaining authorization and approval to use registry data. Ahmad Ishliaq performed the statistical analysis, writing and submitted the manuscript. All authors contributed to the steps of the study design, data collection, analysis and manuscript writing. Myo Nyein Aung finalized the manuscript which was reviewed and approved by all other co-authors. Authors acknowledge the Rachadapisek Sompot Fund for Senior Postdoctoral Fellowship, Chulalongkorn University for supporting this research.

REFERENCES


