

紹介 | Introduction

応用外科解剖学の評価

Assessment of an Applied Surgical Anatomy Module

カーン ファイアーズ アハマト
KHAN Fayyaz Ahmad

尚美学園大学
総合政策学科 講師

Shobi University

2018 年 10 月

Oct.2018

応用外科解剖学の評価

Assessment of an Applied Surgical Anatomy Module

カーン ファイアーズ アハマト
KHAN Fayyaz Ahmad*

[Abstract]

The assessment process in the human anatomy module at most medical schools in Japan use test-centred methods in the form of written examinations. The written assessment process is in Japanese and English and use various standard setting methods. These methods are also used in the applied surgical anatomy module, which is taken as an elective.

日本では、多くの医学部が解剖学の成績評価において筆記試験を用いている。この筆記での評価は日本語と英語で行われており、解剖学での合格基準設定は様々である。この評価方法は選択科目である応用外科解剖学でも用いられる。

English keywords:

Assessment, Applied Surgical Anatomy, Minimally Competent Examinee

* Department of Medical Education, Kyorin University, School of Medicine

1. Introduction

The applied surgical anatomy (ASA) module is taught in English and the students take this module as an elective, which is also known as the student selective component. The students learn ASA in English after passing the human anatomy module assessments in Japanese and English. The method for the summative assessment (assessment of learning) of the written ASA module is selected from the various standard setting methods.

2. Materials and Methods

The two types of standards commonly used are the norm-referenced (relative) and criterion-referenced (absolute)¹⁾. In the norm-referenced standard setting the cut score is set on the number of examinees that will pass, thereby making it relative. The criterion-referenced standard setting shows if the examinee is competent enough to pass or not, and it makes the standard absolute. The latter is a better choice for the ASA module written assessment as it measures the examinee's ability.

The criterion-referenced standard setting is divided into two categories: examinee-centred (EC) and test-centred (TC)¹⁻³⁾. The EC methods focus on the judgement of the examinees' level of performance and the TC methods require a panel of examiners to make judgements based on the assessment items. The TC methods are more suitable for the ASA module assessment process because the required performance level is decided in advance. In the TC methods, four standard setting methods known as the Angoff⁴⁾, Ebel⁵⁾, Nedelsky⁶⁾ and bookmark⁷⁾ are used.

The most frequently used standard setting methods among the criterion-referenced TC methods are Angoff and Ebel (AE). The AE methods determine the passing grade of a test empirically as they have a lot of published research done on them. The AE methods are the most applicable methods for the ASA written assessment process as they rely on the expertise of educators in anatomy and surgery departments. The educators examine the content of each test item and predict how many borderline examinees, also known as the minimally competent examinees (MCE) answer them correctly. In an example of the Angoff method (Fig. 1), five examiners from the anatomy and surgery departments are involved in a six-item test. The examiners are requested to judge each item to determine the probability of the MCE who would answer each item correctly. Each examiner produces a cut score and the overall mean cut score is calculated. In the Ebel Method (Fig. 2), an example of a 30-item test is shown to examiners from the two departments. The examiners consider the content and the level of difficulty of items before judging them to determine the probability of the MCE who would answer them correctly.

Fig. 1 Example of the Angoff Method

Examiner	1	2	3	4	5
Item 1	.80	.75	.80	.65	.70
Item 2	.65	.65	.70	.60	.75
Item 3	.90	.85	.95	.80	.85
Item 4	.75	.80	.75	.70	.65
Item 5	.10	.15	.10	.05	.05
Item 6	.90	.75	.80	.70	.75
Cut score	4.1/6 or 68%	3.95/6 or 66%	4.1/6 or 68%	3.5/6 or 58%	3.75/6 or 63%
Overall Cut score	$4.1 + 3.95 + 4.1 + 3.5 + 3.75 = 19.4/5 = 3.9$ or 65%				

Fig. 2 Example of the Ebel Method

Content	Level of Difficulty		
	Easy	Average	Difficult
Essential	.90 (three items)	.60 (five items)	.30 (four items)
Important	.80 (two items)	.55 (three items)	.20 (three items)
Acceptable	.80 (three items)	.40 (three items)	.15 (four items)
Cut score =	$.90(3) + .60(5) + .30(4) + .80(2) + .55(3) + .20(3) + .80(2) + .40(2) + .15(3) = 14.95/30$ (50%)		

3. Discussion

The Angoff method example (Fig. 1) shows the examiners' cut score ranging from 3.5 (58%) to 4.1 (68%) and the final overall cut score is calculated as 3.9 (65%) to pass the assessment. In the example of Ebel method (Fig. 2), the examiners decide that 3 out of 30 items are essential to the content at the easy level and give the MCE probability of .90. The cut score here is 14.95/30 (50%) to pass the assessment. The panel of educators are trained examiners in the anatomy and surgery departments and they set the pass mark. Although the examiners are expert in the content area involved²⁾, the estimation of the scope of least able examinees who correctly answer each item is a very tiresome task for them. The level of minimal competency is determined when all the educators come upon an agreement. The concept of borderline examinee can be new to some examiners in Japan if it is their first assessment with the AE methods.

4. Conclusion

Since the AE methods are frequently used in high stake assessments, the standard setting process with these methods is a reasonable representation of stakeholder groups. The AE methods are acceptable because some Japanese examiners are familiar with them and they are comparatively easy to use as their application have a considerable amount of published work. Finding an adequate number of educators who have had enough training in using the AE methods might not be easy at a time of globalisation, when more medical subjects are being taught in English. The reliability of the examiners for the ASA module assessment might not be easy and could affect the credibility of the standard⁸⁾.

It is essential to train more examiners to estimate the minimal competency level. This will take time, therefore piloted assessments and trials would be the first step. The piloted summative assessment results can be compared with the Japanese version of the assessment and the formative assessment of the ASA module to rule out any differences due to the language barrier. The results from the AE methods and their relation to the purpose and content of the assessment can also be compared, since no research has identified one method being better in all assessments^{1, 9)}.

References

1. Livingston, S.A. And Zieky, M.J: Passing Scores: A manual for setting standards of performance on education and occupational tests. Princeton, NJ; Educational Testing Service, 1982. p. 1-71
2. Jaeger, R.M: Certification of student competence. In R.L. Linn (ed.) Educational Measurement. New York: American Council on Educational and Macmillan, 1989. p. 485-514
3. Hambleton, R.K., Jaeger, R.M., Plake, B.S., and Mills, C: Setting performance standards on complex educational assessments. *Appl Psychol Measurement*. 23(4): 355-366, 2000
4. Angoff W.H: Scales, norms, and equivalent scores. In: Thorndike RC, editor. Educational measurement. 2d ed. Washington, DC, American Council on Education, 1971. p. 508-600.
5. Ebel, R. L: The practical validation of tests of ability. *Educational Measurement: Issues and Practice*, 2: 7-10, 1983
6. Nedelsky, I: Absolute grading standards for objective tests. *Educ Psychol Measurement*. 14(1):3-

19, 1954

7. Lewis, D. M., Mitzel, H. C., & Green, D. R: Standard setting: A bookmark approach. In D. R. Green (Chair), IRT-based standard-setting procedures utilizing behavioural anchoring. Symposium conducted at the Council of Chief State School Officers National Conference on Large-scale Assessment, Phoenix, AZ, 1996
8. Verhoeven, B. H., Verwijnen, G. M., Muijtjens, A. M., Scherpbier, A. J., & van der Vleuten, C. P: Panel expertise for an Angoff standard setting procedure in progress testing: item writers compared to recently graduated students. *Med Educ*, 36(9), 860-867, 2002
9. Downing, S. M., Tekian, A., & Yudkowsky, R: Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med*, 18(1), 50-57, 2006