

## Diagnostic accuracy study of internal assessment using receiver operator characteristic curve analysis

Khadija Qamar, Muhammad Alamgir Khan, Sadaf Saleem

### Abstract

**Objective:** To assess whether internal assessment as a part of continuous assessment links to the outcome of the final summative assessment.

**Methods:** The diagnostic accuracy study was conducted at Army Medical College, Rawalpindi, Pakistan, from November 2015 to July 2016, and comprised medical students of 2nd year. Different teaching methods used were interactive lectures, case-based sessions, demonstrations, small group discussions, skill lab and practicals. Other confounding factors were not considered. Receiver operator characteristic curve was computed to determine diagnostic accuracy of internal assessment for the prediction of examination results.

**Results:** Out of 202 students, 122 (60.4%) were male and 80 (39.6%) were female with an overall mean age of  $20.05 \pm 0.69$  years. Total marks of 2nd professional examination and internal assessment were normally distributed with mean values of  $131.71 \pm 19.81$  and  $36.18 \pm 8.03$  respectively. The cut-off value was 27.5 and at this value, sensitivity was 100% and specificity was 91%.

**Conclusion:** Diagnostic power of internal assessment to identify students who may fail in professional examination was significantly high.

**Keywords:** Curriculum, Internal assessment, Medical education, Anatomy. (JPMA 68: 721; 2018)

### Introduction

Assessment is the systematic process of documenting knowledge, skills, attitudes and beliefs, review and use of this information to improve student learning.<sup>1</sup> Continuous assessment is the process of pursuing and interpreting evidence that is to be used by learners and their teachers to decide what the level of learner is, what they need to do to achieve better scores and how to do it. The purpose of continuous assessment plan is to monitor the learning, identifying student's strengths and weaknesses, assuring achievement of predetermined learning objectives and identify the failures.<sup>2</sup> Therefore it can provide prompt pointers on students' performance.<sup>3</sup>

The causes of academic failure can be varied. The issues may be classified as academic and personal issues/difficulties. The causes of poor academic results may widely vary, but the subjects with which candidates face difficulty are usually similar. Current literature has reported that students have a generally preferred learning style, but will adopt their way of learning to their concept of what is required of them.<sup>4</sup>

It is a part of the teaching process of a teacher to observe learning and this can be conveniently done through internal assessment. Hence, informed decisions about

students' learning styles and needs can be based on statistics.<sup>5</sup> These can also help students in their self-assessment and provide them with effective in-time feedback. Special tasks are designed for a particular concepts or skill and this can be made a part of internal assessment process which can provide the teacher valuable information about student's performance. Particularly useful examples of internal assessment are class assignments, term tests and viva or objective structured practical examination (OSPE).

Internal assessment was introduced and implemented in the medical colleges so that students study throughout the year as they will have an impetus to study: To obtain good grades. Initially, the weightage of the internal assessment was low, and most of the time it hardly had an influence on the final outcome. However, with the passage of time, as the weightage of the internal assessment was increased, it became pivotal in deciding who fails and who passes the final professional examination. A good internal assessment also gives a fair idea to the examiner about how serious the student was throughout the year.<sup>6</sup>

The explanation behind the accuracy of internal assessment in predicting who is more likely to fail in the final examination is quite plausible. It is most likely that those students who perform poorly throughout the year will also perform poorly in their final examination,

.....  
Army Medical College, National University of Medical Sciences, Rawalpindi.

**Correspondence:** Sadaf Saleem. Email: sadafsaleem15875@gmail.com

because they are unable to cover the vast course in a matter of few weeks. Unfortunately, many students are under the false impression that they can cover an entire year's course in a matter of few weeks just to realise later that the course is just too much to study. But by that time it is already too late. The student then ends up doing selective studies, which is a high risk to take, not to mention that it is morally wrong as well. Moreover, the realisation that the course is too much leads to anxiety attacks, depression and lack of concentration, which all contribute to the student performing even more poorly.

In order to ensure that the results produced by students are authentic and in line with the requirements of the conferring university/medical colleges, these assessments are committed to a quality assurance process of internal assessment and standardisation. Internal assessment is based on the results of all the tests that are part of continuous assessment throughout the academic session. Therefore, in order to enhance the quality and accountability of medical education, designing a valid and reliable assessment system is essential.<sup>7</sup>

In order to improve the outcomes for low-achieving students, early identification and remediation are critical to educators' success. A number of studies have considered the relationships between various teaching methods and medical students' academic performance.<sup>8</sup> However, very few studies have been done to examine the relationships between students' performance in final exams and their internal assessment marks. The present study was planned to determine the degree to which the internal assessment correlated with prediction of final outcomes in professional examination.

### Subjects and Methods

The diagnostic accuracy study was conducted at Army Medical College, Rawalpindi, Pakistan, from November 2015 to July 2016, and comprised medical students of 2nd year. After taking approval from institutional ethics committee, sample size was calculated using Buderer's formula.<sup>9</sup> By keeping sensitivity level at 0.98, expected specificity at 0.90, expected prevalence at 0.05, absolute precision at 0.1 and confidence level at 95%, a sample size of 161 was calculated. However, the whole class of 2nd year

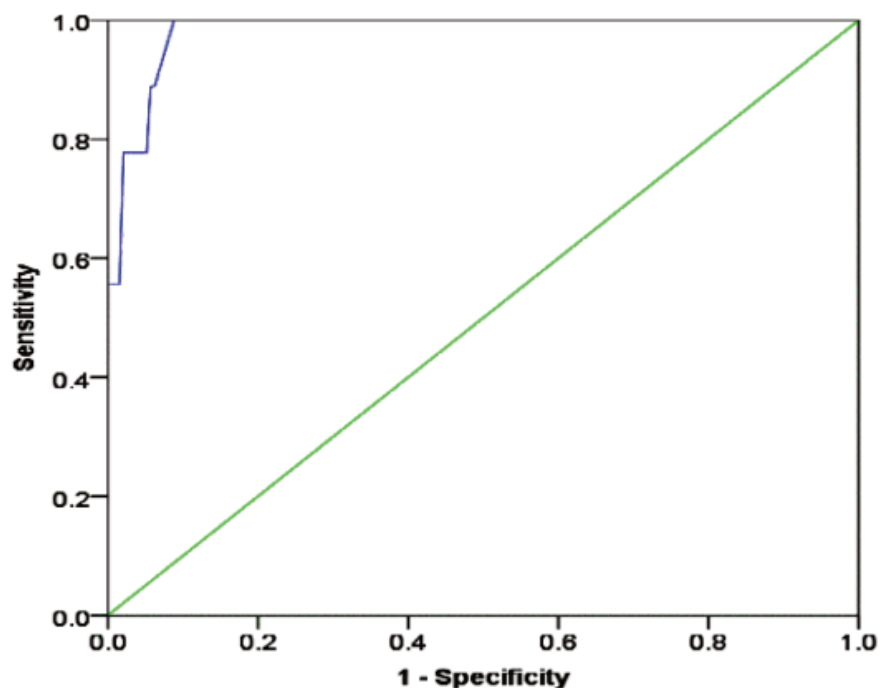
MBBS was included in the study.

Internal assessment of 2nd year MBBS class and the result of Anatomy paper of the same class in 2nd Professional MBBS examination were utilised for analysis. Total marks for the paper were 200 which were further divided into 100 marks for both theory and practical exams. Students with less than 50% marks in either component were declared fail and equal to or above 50% were declared pass. A total of 60 marks were assigned to internal assessment with a division of 30 each to theory and practical. The results of 3 modular and 1 pre-annual examination constituted the internal assessment.

SPSS 23 was used for data analysis. Numerical variables like age, marks in professional exam and internal assessment were used to calculate mean and standard deviation. Categorical variables like gender and result of professional examination were used to work out the frequency and percentage. Receiver operator characteristic (ROC) curve was computed to determine diagnostic accuracy of internal assessment for the prediction of examination results. Alpha value was kept at 0.05.

### Results

Out of 202 students, 122 (60.4%) were male and 80 (39.6%) were female with an overall mean age of  $20.05 \pm 0.69$  years. Normal distribution of 2nd professional examination and



**Figure:** Receiver Operator Characteristic Curve for internal assessment and result of 2nd Professional MBBS Examination result.

**Table-1:** Area under curve along with other statistical indicators for Receiver Operator Characteristic Curve.

Area under curve	Standard error	p-value	95% confidence interval	
			Lower bound	Upper bound
0.98	0.01	<0.001*	0.96	1.00

\*p-value significant (<0.05).

**Table-2:** Various cut-off values of internal assessment at different sensitivity and specificity.

Fail in Prof if less than or equal to	Sensitivity	1-Specificity
14.00	0.000	0.000
15.50	0.111	0.000
17.00	0.333	0.000
18.50	0.556	0.000
19.50	0.556	0.005
20.50	0.556	0.016
21.50	0.778	0.021
22.50	0.778	0.026
23.50	0.778	0.041
24.50	0.778	0.052
25.50	0.889	0.057
26.50	0.889	0.062
27.50*	1.000 (100%)	0.088 (91% specificity)
28.50	1.000	0.130
29.50	1.000	0.150
30.50	1.000	0.197
31.50	1.000	0.244
32.50	1.000	0.290
33.50	1.000	0.337
34.50	1.000	0.363
35.50	1.000	0.451
36.50	1.000	0.503
37.50	1.000	0.528
38.15	1.000	0.570
38.65	1.000	0.575
39.50	1.000	0.627
40.50	1.000	0.684
41.50	1.000	0.731
42.50	1.000	0.746
43.50	1.000	0.793
44.50	1.000	0.813
45.50	1.000	0.850
46.50	1.000	0.876
47.50	1.000	0.912
48.50	1.000	0.943
49.50	1.000	0.969
50.50	1.000	0.979
51.50	1.000	0.995
53.00	1.000	1.000

\*Cut off value at maximum sensitivity and specificity.

internal assessment marks was observed with mean values of 131.71±19.81 and 36.18±8.03 respectively.

ROC curve and area under the curve were also figured out (Figure; Table-1).

The cut-off value of internal assessment, at which both sensitivity and specificity were maximum, was 27.50. At this cut-off value, sensitivity was 100% and specificity was 91% (Table-2).

### Discussion

Progressive in-course assessments have become a useful tool in monitoring and evaluating students' understanding and grasp of a subject. The performance of a student of 2nd Professional MBBS course in the subject of Anatomy throughout the specified session was projected through internal assessment results and was supposed to be a forecast for the final summative examination result. We observed that the diagnostic power of internal assessment to identify students who may fail in professional examination was significantly high. In a similar study done at 6 schools in England, the teachers were asked to make specific formative assessments for the students. There was ample evidence that improving formative assessment does produce tangible benefits in terms of externally mandated assessments (such as key stage 3 tests and General Certificate of Secondary Education [GCSE] examinations in England).<sup>10</sup> Improvement of performance after going through the formative assessments has been suggested in other studies<sup>11-13</sup> as well. Here the effects of formative assessments have been studied. The factors considered were inferential, correlational analyses based on linear causal framework. Another study done on first year medical students in Canada investigated the effects of two formative assessments and their relation to the summative assessment.<sup>14</sup> It was concluded that summative assessments can be predicted by formative assessments in medical students.

Although continuous assessment constitutes a fundamental part in the curriculum, the 'pass' and the 'fail' certificates are based on students' performance in the final summative examination. Importantly, assessment

results should provide students with meaningful feedback on their strengths and weaknesses. Similarly, assessment results should provide useful feedback to teachers and future employers. Although students can escape from poor teaching by independent learning, they cannot escape from the effects of poor assessments as they have to pass the examination.<sup>13</sup> One must consider some essential factors when designing assessments as discussed below.

Internal assessment marks system gives teachers a chance to know students' knowledge both in theory and practical. Internal assessment can compensate for many of the drawbacks of the year end examination and enrich the assessment process. The reluctance to exploit the potential of internal assessment is related largely to unawareness regarding its several aspects. When properly implemented, internal assessment is better than the year end examination in terms of its validity, reliability (consistency of performance), feasibility and educational impact. To ensure that students are not denied the benefit of this extremely useful modality, efforts need to be made to improve its implementation and acceptability. The effectiveness of continuous assessment was emphasised in virtually all educational settings and levels of education. Also noted was the fact that grades and marks could be ineffective in providing constructive feedback, especially to the low-ability learners.<sup>10</sup>

A meta-analysis of meta-analytic studies indicated that the single most important factor in promoting learning is feedback. Feedback should be offered to students while they still have the chance to improve. This will help the students to come up with metacognitive strategies to overcome their deficiencies.<sup>15</sup> It should be the primary concern, especially of internal assessment, that the results should be utilised for the improvement of the quality and quantity of learning, eventually affecting the summative assessment.

### Conclusion

Diagnostic power of internal assessment to identify students who may fail in professional examination was significantly high. Internal assessment at cut-off value where sensitivity was 100% along with maximum possible specificity may act as a screening test to isolate the students who are at potential risk of failing the

professional examination. This would form a subset of students for whom special coaching classes or counselling sessions shall be arranged so that the predictable failure in professional examination can be avoided.

**Disclaimer:** None.

**Conflict of Interest:** None.

**Funding Disclosure:** None.

### References

1. Vleuten C Van Der. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996; 1: 41-67.
2. Fowell S, Southgate L. Evaluating assessment: the missing link? *Med Educ.* 1999; 33: 276-81.
3. Chew B, Zain A, Hassan F. Emotional intelligence and academic performance in first and final year medical students: a cross-sectional study. *BMC Med Edu.* 2013; 27: 13-44.
4. Harden R. AMEE Guide No. 21: Curriculum mapping: a tool for transparent and authentic teaching and learning. *Med Teach.* 2001; 23: 123-37.
5. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach.* 2012; 34: 960-92.
6. Cook D, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Acad Med.* 2016; 91: 785-95.
7. Johnson TR, Khalil MK, Pepler RD, Davey DD, Kibble JD. Use of the NBME Comprehensive Basic Science Examination as a progress test in the preclerkship curriculum of a new medical school. *Adv Physiol Educ.* 2014; 38: 315-20.
8. David M, Davis M, Harden R, Howie P. AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment. *Med Teach.* 2001; 23: 535-51.
9. Naing L. Sample size calculation for sensitivity and specificity studies. 2004
10. Barbarà-i-Molinero A, Barbarà-i-Molinero A, Cascón-Pereira R, Cascón-Pereira R, Hernández-Lara AB, Hernández-Lara AB. Professional identity development in higher education: influencing factors. *Inter J Educ Manag.* 2017; 31: 189-203.
11. Bondemark L, Knutsson K, Brown G. A self-directed summative examination in problem-based learning in dentistry: a new approach. *Med Teach.* 2004; 26: 46-51.
12. Greer L. Does changing the method of assessment of a module improve the performance of a student? *Ass Eval Higher Educ.* 2001; 26: 127-38.
13. Thissen-Roe A, Hunt E, Minstrell J. The DIAGNOSER project: Combining assessment and learning. *Behav Res Meth.* 2004; 36: 234-40.
14. Krasne S, Wimmers PF, Relan A, Drake TA. Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Adv Health Sci Educ.* 2006; 11: 155-71.
15. Rolfe I, McPherson J. Formative assessment: how am I doing? *Lancet.* 1995; 345: 837-9.