ISSN: 2091-2749 (Print)
2091-2757 (Online)

Evaluation of web-based self-assessment module administered in a medical school of Nepal

Shital Bhandary,[1] Satish Raj Ghimire,[2] Ira Shrestha[3]

Assistant Professor
Department of Community Health Sciences,[1] Department of Anatomy,[2]
Department of Physiology,[3] Patan Academy of Health Sciences, Lalitpur, Nepal

Correspondence
Shital Bhandary
Coordinator, Examination Management Committee,
Patan Academy of Health Sciences,
Lalitpur, Nepal
E-mail: shitalbhandary@pahs.edu.np
Phone: +977-1-5545112

Peer reviewed by
Prof. Dr. Jay N Shah
Patan Academy of Health Sciences

Dr. Ashis Shrestha
Patan Academy of Health Sciences

ABSTRACT

Introductions: Self-assessment enables medical students to self-evaluate their knowledge and seek timely assistance for effective learning from their peers and faculties. Self-assessment is an integral part of the student assessment system of School of Medicine, Patan Academy of Health Sciences.

Methods: Multiple Choice Questions (MCQs) and Structured Integrated Short Answer Questions (SISAQs) links were sent to the personal e-mail of medical students at the start of Principle of Human Biology I Block. These items were created and selected from various disciplines as per their curricular weightage in this block. Students' scores and feedback were analysed once this 11-week long block was over.

Results: The MCQ surveys had ideal difficulty levels and acceptable discrimination indices but they had poor internal consistencies. The criterion-referenced cut-scores of MCQs and SISAQ were higher than the conventionally used 50% pass-mark in Nepal. Students' suggested to increase the MCQs and SISAQ numbers and match them in terms of their difficulty levels with the end-block and end-year summative assessments. Item analysis helped to identify the items to be retained, revised and discarded for the future use.

Conclusions: Web-based self-assessment of knowledge was found to be an extremely useful tool to inculcate self-directed and life-long learning habits among medical students.

Keywords: self-assessment, MCQ, SISAQ, angoff, borderline regression, standard setting

Plain Language Summary

Self-assessment of knowledge was administered to the medical students in a medical school in Nepal. It helped them to self-evaluate their knowledge and seek timely assistance from their peers and proper guidance from their faculties.

INTRODUCTIONS

Self-assessment is essential for self-directed learning and practice of medicine.[1] Self-assessment enables students to self-evaluate their knowledge and skills before and during any course and seek timely assistance for effective and life-long learning.

Self-directed learning is the main teaching/learning strategy of undergraduate medical education program at School of Medicine, Patan Academy of Health Sciences (PAHS-SOM) as its courses heavily use Problem Based Learning (PBL).[2] Thus, self-assessment of knowledge is also an integral part of student assessment system to help the students to appraise depth and breadth of knowledge *a priori*. Consequently, Multiple Choice Questions (MCQs) and Problem Based Questions (PBQ) are provided to them so that they can self-evaluate their knowledge during the organ-system blocks of the hybrid PBL program.

This study aims to evaluate the self-assessment module implemented in one of the organ system block of integrated basic sciences phase of undergraduate medical curriculum at PAHS-SOM.

METHODS

Self-assessment question for the 11 weeks long (Nov 2010-Feb 2011) Principle of Human Biology I (PHB I) Block consisted of 30 A-Type MCQs.[3] and one PBQ. Each MCQ carried one mark where as the PBQ carried 15 marks. There were no negative markings in these items. The self-assessment was designed for the 58 pioneer batch students of undergraduate medical education program of PAHS-SOM. The MCQs were constructed using appropriate clinical, lab and public health vignettes[3] to test progressively higher level of knowledge acquisition.[4]

Six basic sciences subjects (Anatomy, Biochemistry, Microbiology, Pathology, Pharmacology and Physiology) along with Introductory Clinical Medicine (ICM) or Early Clinical Exposure] and Community Health Sciences (CHS) contributed MCQs as per the curriculum blueprint, which was based on their planned contents on PBL and other didactic sessions of this block. Each MCQ was discussed in the presence of at least six trained faculty members of different disciplines to ensure its face and content validity[5] followed by standard setting process using a criterion-referenced (Modified Angoff) method, which established the reliable pass mark for the test *a priori*.[6]

The PBQ was intended as an integrated content assessment of the PBL sessions of PHB I block. It was developed in a half-day long PBQ writing workshop where all the faculty involved in the block were present. At first, discipline-wise topics was listed followed by selection of a suitable clinical scenario (e.g. pleural efusion) for the PBQ in consensus. The PBQ consisted of vignettes and questions from various disciplines along with specific model answers and clear marking schemes.

Thus, an innovative Structured Integrated Short Answer Question (SISAQ) was developed to assess Higher Order Thinking Skills (HOTS) among medical students at PAHS-SOM. This SISAQ was standard set using a five ordinal categories (Very Poor, Poor, Borderline, Good and Very Good) based on its clinical importance and difficulty levels of the questions. Each categories represented "range of marks" agreed in consensus. A criterion-referenced (borderline regression) method was used to calculate its cut score (pass mark) after the test.[5,6]

The MCQs and SISAQ were designed in the web-based Survey Monkey software using its basic account. As basic account allowed only 10 questions in a survey, four surveys were created. First three surveys contained 10 MCQs each whereas fourth survey contained one SISAQ with eight items. The survey links were sent through the personal e-mail address of the students and they were available throughout the PHB I block. Since these surveys were anonymous, personal and other details of the students were not collected.

After the PHB I block was over, students' responses were copied to the Microsoft Excel spreadsheet. MCQs were scored using its "keyed" responses whereas SISAQs were scored using model answers and it's marking schemes. Test and item analysis,[7,8] of MCQs and SISAQ were done using Microsoft Excel 2013. Students' feedback were analysed using manual content analysis. Ethical approval for the study was granted by the Institutional Review Committee, PAHS.

RESULTS

Web-based MCQ and SISAQ administration went well without technical difficulty. Out of 58 students, MCQ part I, II and III were completed by 47, 45 and 43 respectively whereas SISAQ was done by 37. (Table 1.)

Table 1. Test analysis of web based self assessment MCQs to medical students.

MCQ	N	Items	Mean	SD	Alpha	SEM	Difficulty Index	Discrimination Index
Part I	47	10	7.00	1.49	0.49	1.07	0.72	0.32
Part II	45	10	6.36	1.19	0.56	0.79	0.65	0.19
Part III	43	10	5.35	1.38	0.14	1.28	0.54	0.30

Note: SD = Standard Deviation, SEM = Standard Error of Measurement, Alpha = Cronbach's alpha, MCQ = Multiple Choice Question

Test analysis of the MCQs revealed increasing difficulty levels for Part I, Part II and Part III as mean scores decreased respectively for these parts. Score variation was high for Part I followed by Part III and Part I as per standard deviation of scores. The internal consistency reliability was higher for part II than part I and III. SEM values indicated a wide confidence intervals for Part III and Part I than Part II. MCQs were of ideal difficulty (0.3-0.7) and/or ideal discrimination (>=0.25) levels for Part I and III. MCQ Part II had ideal difficulty level for 4-option MCQ but it had marginal discrimination (0.15-0.24) level among high and low performing students. (Table 1.)

Around 76, 34 and 21 percent of students passed in the MCQ part I, II and III as the criterion-referenced pass percentage were 66.8%, 67.2% and 61.8% respectively. Similarly, the cut-off scores for the Basic Sciences, CHS and ICM MCQs were 67.2%, 63.9% and 59.5% respectively.

SISAQ cut-off score was 8.51 or 56.7%. This criterion-referenced pass-mark was also higher than the conventional fixed standard of 50% but it was lower than the pass-marks of all 3 MCQs surveys. (Fig. 1.)

Table 2. Cut off score and outcome analysis of web based self assessment MCQs to medical students

MCQ	Items	Cut-off score* Mean (%)	Pass N (%)	Fail N (%)	Discipline		Items (%)	Cut-off Score* Mean (%)
Part I	10	6.68 (66.8)	35 (74.5)	12 (25.5)	Basic Sciences		18 (60)	6.72 (67.2)
Part II	10	6.72 (67.2)	16 (34.0)	31 (66.0)	Community Health		8 (27)	6.39 (63.9)
Part II	10	6.18 (61.8)	9 (20.9)	34 (79.1)	Introductory Medicine	Clinical	4 (13)	5.95 (59.5)

* Pass Marks: Obtained from criterion-referenced standard setting (Modified Angoff) method.[5,6]

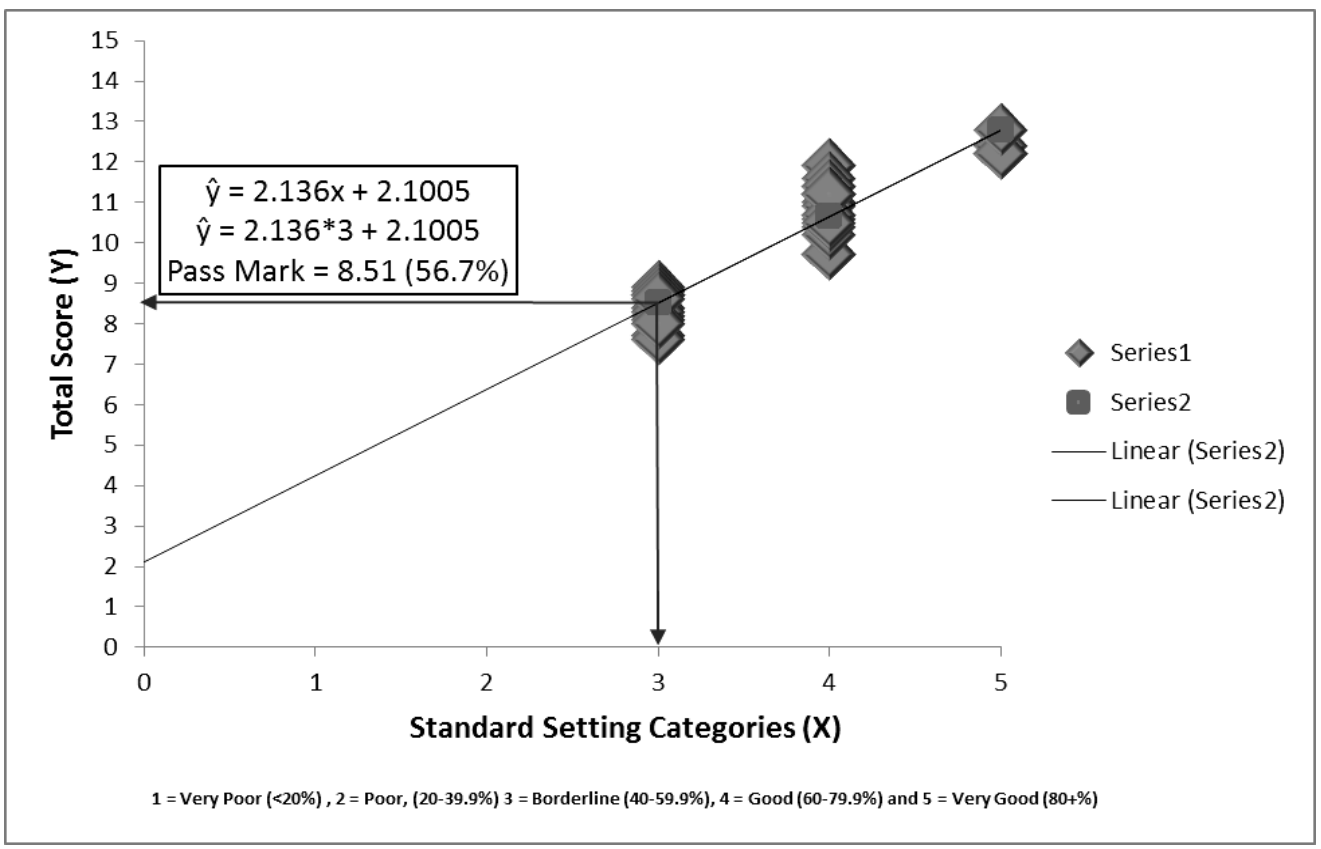

Figure 1. Cut-off score by borderline regression method, to medical students web based self assessment SISAQ.

Table 3. Test Analysis and Outcome of Self-Assessment SISAQ

Group	N	Items	Mean (%)	SD (%)	Discrimination	Reliability	Pass (%)	Fail (%)
Low	10	8	8.30 (55.3)	0.48 (3.78)	NA	NA	3 (30)	7 (70)
Mid	17	8	10.41 (69.4)	0.71 (3.22)			17 (100)	0 (0)
High	10	8	11.90 (79.3)	0.57 (4.75)			10 (100)	0 (0)
Total	37	8	10.24 (68.3)	1.48 (9.86)	0.24	0.14	30 (81.8)	7 (18.9)

Note: NA = Not Applicable, Low = Below 27th percentile and High = Above 73rd percentile of total score, SISAQ: Structured integrated short answer question

On average, students scored 68.3% in the SISAQ where low, mid and high performing group scored 55.3%, 69.4% and 79.3% respectively. The discrimination index was within acceptable range but the internal construct reliability was very poor. Standard deviations between low, mid and high groups point out discrimination within these groups too. Criterion-referenced (borderline regression method,[5,6]) cut-score categorized around 82% of the students as "pass" and about 19% as "fail" in the SISAQ. (Table 3)

Quality Assurance (QA) indices for MCQs and SISAQ using Classical Item Analysis[7,8] showed all students answered two MCQs (QN6 and QN12) correctly and thus had difficulty index of one and discrimination index of zero. Similarly, two MCQs (QN15 and QN16) were answered incorrectly by all the students and thus both difficulty and discrimination indices were zero. Five MCQs (QN11, QN15, QN24, QN30 and QN3) had ideal difficulty (0.3-0.7) as well as excellent discrimination (>=0.35) indices whereas three MCQs (QN4, QN11 and QN29) had ideal difficulty with good discrimination indices (0.25 – 0.34). In addition, all the four options were selected for three MCQs (DE=100%) only whereas only 1, 2 and 3 option/s was/were selected for 10 MCQs (DE=25%), seven MCQs (DE=50%) and 10 MCQs (DE=75%) respectively. (Table 4)

Out of eight items used in the SISAQ, one item (QN31) had ideal difficulty (0.3-0.7) and good discrimination (0.25-0.34) indices. Similarly, two items (QN34 and QN36) were of low difficulty (>=0.7) but had excellent discrimination (>=0.35) indices whereas one item (QN38) had ideal difficulty with marginal discrimination (0.15-0.24) indices. On the other hand, two items (QN7 and QN33) had ideal difficulty but poor discrimination (<0.15) indices. Further, QN35 was easy (difficulty index > 0.8) and could not discriminate the high and low performers (discrimination index = 0.00) whereas QN32 was confusing to the high performers than low performers as its discrimination index was negative. Item Reliability Indices were above 0.4 for QN4 and QN6 whereas it was below 0.25 for other items suggesting problems with the item and the total scores.

Table 4: Classical item analysis of self assessment MCQs and SSIAQ to medial students

MCQ	Part I										Grand Mean
Item	QN1	QN2	QN3	QN4	QN5	QN6	QN7	QN8	QN9	QN10	
Difficulty	0.21	0.83	1.00	0.59	0.91	1.00	0.53	0.93	0.29	0.89	0.72
Discrimination	0.52	0.50	0.08	0.30	0.33	0.08	0.13	0.42	0.52	0.27	0.32
DE[a]	50%	75%	25%	100%	50%	25%	75%	50%	75%	50%	58%
Discipline	Anatomy			Physiology			Biochemistry			Micro	
MCQ	Part II										Grand Mean
Item	QN11	QN12	QN13	QN14	QN15	QN16	QN17	QN18	QN19	QN20	
Difficulty	0.36	1.00	0.00	0.95	0.43	0.00	0.86	0.98	0.98	0.98	0.65
Discrimination	0.49	0.00	0.00	0.15	0.90	0.00	0.13	0.05	0.10	0.10	0.19
DE[a]	75%	25%	25%	25%	50%	25%	75%	25%	25%	25%	38%
Discipline	Microbiology		Pharmacology			Pathology			ICM		
MCQ	Part III										Grand Mean
Item	QN21	QN22	QN23	QN24	QN25	QN26	QN27	QN28	QN29	QN30	
Difficulty	0.70	0.44	0.81	0.60	0.77	0.98	0.09	0.02	0.49	0.47	0.54
Discrimination	0.29	0.18	0.64	0.70	0.20	0.09	-0.01	0.04	0.31	0.58	0.30
DE[a]	75%	100%	75%	75%	50%	25%	75%	75%	50%	100%	70%
Discipline	ICM		Community Health Sciences								
SISAQ	Structured Integrated Short Answer Question (SISAQ)										Grand Mean
Item	QN31	QN32	QN33	QN34	QN35	QN36	QN37	QN38	-	-	
Difficulty	0.69	0.65	0.35	0.78	0.85	0.72	0.82	0.51	-	-	0.68
Discrimination	0.25	-0.05	0.10	0.55	0.00	0.70	0.25	0.20	-	-	0.24
IRI[b]	0.07	0.02	0.14	0.49	0.07	0.43	0.24	0.06	-	-	0.14
Discipline	ICM	Pathology			Physiology		Pharmacology	Anatomy	-	-	

Note: a. Distractor Efficiency, b. Item Reliability Index

DISCUSSIONS

The self-assessment of knowledge using MCQs and SISAQ was highly appreciated by the pioneer batch of medical students of PAHS-SOM as they were able to practice the questions during the course and realize the changes in their understanding and knowledge application as the course moved forward. They also informed that they "pursued help from their peers and/or faculty whenever required to solve the items correctly and/or seek clarifications on the questions". The keys of the MCQs and model answers of SISAQ shared after the completion of the PHB I block gave them chance to correct their mistakes and seek the further help from peers or faculty for the end-block examination.

As the Cronbach's alpha were less than 0.60 for all parts of MCQs, low correlation between the MCQs were indicated. Nonetheless, overall alpha could reach to 0.64 and 0.70 when the test length was increased by four-fold (i.e. 40 MCQs) and five-fold (i.e. 50 MCQs) respectively.[9] This suggested that at least 50 MCQs were required to achieve the minimum accepted internal consistency reliability of 0.70 even for self-assessment test with valid responses from 45 students or more in each organ-system blocks of PAHS-SOM.[2]

Most of the students' also advised to increase self-assessment MCQs and PBQs numbers and suggested to match the questions with that of the continuous end-block assessments as they found "more memory level MCQs in the self-assessment than in the end-block assessment". Based on this feedback, examination section was able to share 50 self-assessment MCQs and three self-assessment SISAQs for subsequent organ-system blocks with timely assistance from all the faculty involved in the integrated basic sciences program. Further, faculty revised/added substantial number of MCQs assessing higher level knowledge of Bloom's Taxonomy.[3,4] in all the self-assessments. This ensured reliability and validity of the self-assessment module and items included for each organ-system blocks at PAHS-SOM.

Students' feedback further revealed SISAQ vignette and questions being "difficult to understand". This prompted the moderation of all the SISAQs constructed for a particular block by a panel of experts, which ensured the SISAQs being easier to understand irrespective of it being selected for the formative self-assessment or continuous end-block assessments.

The test-based criterion-referenced standard setting (Modifed Angoff) [5,6] cut-score was 65% for these 30 MCQs to pass the self-assessment test, which was higher than the conventional fixed standard of 50% in Nepal and South Asia.[5] Further, as these MCQs were constructed to assess the comprehension and application level of Bloom's taxonomy,[3,4] their cut-off score should have been lower than the memory level MCQs.[5,10,11] Careful analysis of these MCQs revealed higher Angoff scores even for the MCQs assessing higher knowledge levels, which is partly due to the emphasis on the achievement of the competencies and/or problem in defining "borderline students" during the standard setting processes.[5,6]

The result were discussed with the faculty and they started providing plausible and defensible cut-scores for old and new MCQs in the subsequent standard setting sessions. Further, it was also recommended to adjust the cut-scores of all the end-block assessments in the basic sciences using compromised methods such as Hofstee or Beuk to compensate the upward bias and errors.[5,6,11]

The exam-centered criterion-standard setting method (Borderline Regression method),[5,6,11] used in the SISAQ revealed a cut-score of around 57%, which is found to be a plausible and defensible one as it had quite a few memory level questions despite being constructed to assess comprehension, application and analysis levels of Bloom's Taxonomy.[3] Yet, SISAQ was clearly able to discriminate the low, mid and high performers. Similarly, Borderline Regression Method automatically adjusted the cut-score of the SISAQ based on student's actual score against the criteria set by the faculty. However, the main problem of the self-assessment SISAQ used was low item reliabilities which was partly due of the low response rate of the students. Item reliabilities increased substantially when 3 self-assessment SISAQs were sent to the students in the subsequent organ-system blocks.

Item analysis,[7] of MCQs and SISAQ gave insight on the item/s that were working, needs revision and/or to be discarded for the future use,[8] which in turn provided the best opportunity to explain the importance of item analysis as the quality assurance tool for the student assessment system at PAHS.[8,10] In addition, item Analysis of PHB I end-block assessment was presented to the faculty and academic leaders of PAHS and it was accepted as the main tool for quality assurance and question banking system of PAHS-SOM.

As the self-assessment was anonymous, its effect on the end-block assessment could not be assessed at individual level, which was the main limitation of this study. However, feedback of the students and the result of the end-block formative assessment clearly showed the importance of self-assessment of knowledge for the hybrid PBL curriculum at PAHS-SOM.

CONCLUSIONS

Web-based self-assessment of knowledge enabled medical students to continuously self-evaluate their knowledge and motivated them to seek help from peers and guidance from faculty whenever required.

ACKNOWLEDGEMENTS

We would like to thank all the pioneer batch medical students of PAHS-SOM. We express our sincere gratitude to Professor Dr Rajesh Nath Gongal, Founding Dean and all the basic sciences faculty involved in the first PHB I block of PAHS-SOM for their timely and prompt assistance during this study.

REFERENCES

1. Antonelli M. Accuracy of second-year medical students' self-assessment of clinical skills. Academic Medicine. 1997;72(Supplement 1):S63-S65.
2. Ghimire SR, Bhandary S. Student's perception and preference of problem based learning during introductory course of a nepalese medical school. Journal of Patan Academy of Health Sciences. 2002;1(1):64-68.
3. Case S, Swanson D. Constructing written test questions for the basic and clinical sciences, national board of medical examiners, 3rd Edition [Internet]. 2002 [cited 5 July 2015]. Available from: http://www.nbme.org/publications/item-writing-manual.html
4. Carneson J, Delpierre G, Masters K. Designing and Managing MCQs: Appendix. C: MCQs and Bloom's Taxonomy [Internet]. 2011 [cited 5 July 2015]. Available from: http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html
5. Bhandary S. Standard Setting in Health Professions Education. Kathmandu Univ Med J. 2011 Jan-March;9(33):3-4.
6. Yudkowsky R, Downing SM, Tekian A. Standard Setting. In: Downing SM, Yudkowsky R, editors. Assessment in Health Professions Education. New York, Routledge; 2009.110-48p.
7. Hingorjo MR, Jaleel F. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. J Pak Med Assoc. 2012;62(2):142-7.
8. Downing SM. Statistics of Testing. In: Downing SM, Yudkowsky R, editors. Assessment in Health Professions Education: New York, Routledge. 2009.119-48p.
9. Axelson RD, Kreiter CD. Reliability. In: Downing SM, Yudkowsky R, editors. Assessment in Health Professions Education: New York, Routledge. 2009.57-74p.
10. Haldayna TM. Developing and Validating Multiple-Choice Test Items. 3rd ed. New Jersey: Lawrence Erlbaum Associates; 2004.
11. Cizek GJ. Standard Setting. In: Downing SM, Haladyna TM, editors. Handbook of Test Development. New Jersey; Lawrence Erlbaum Associates; 2006.225-58p.