นิพนธ์ต้นฉบับ

Quality of Multiple Choice Question tests, Faculty of Medicine, Chulalongkorn University: first semester of academic year 1993.*

Boonnart Laisnitsarekul**

Laisnitsarekul B.Quality of Mutiple Choice Question tests, Faculty of Medicine, Chulalongkorn University : first semester of academic year 1993. Chula Med J 1994 Jul;38(7):391-398

This descriptive research studied the quality of MCQ tests used during the first semester of academic year 1993, administered by the Faculty of Medicine, Chulalongkorn University. Assessed were item-difficulty index, item-discrimination index, reliability of tests and identification of the number of questions rated an excellent, good, revise, poor and negative by using standard criteria of the World Health Organization. Thirty-two MCQ tests were analysed by the CTIA item analysis program. There were 2,866 items: 2,703 items (94.31%) of the "one best answer" type and 163 items (5.69%) of the k type. Reliability of the tests studied was in the range of 0.132-0.914. Nineteen tests (59.38%) had a reliability indicator equal to 0.70 or more and 13 tests (40.63%), the reliability indicator was below 0.70. Out of 2,866 questions, 480 were rated as excellent questions (16.75%), 219 good (7.64%), 525 revise (18.32%), 1,403 poor (48.95%) and 239 negative (8.34%). Excellent and good questions would therefore be most useful in preparing a question bank.

Key words : *MCQ, Difficulty Index, Discrimination Index, Reliability, Question Bank, Test.*

Reprint request : Laisnitsarekul B. Medical Education Unit, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand.

Received for publication. April 5, 1994

* The Rachadapiseksompoj-China Medical Board Research Funds.

** Medical Education Unit, Faculty of Medicine, Chulalongkorn University.

บุญนาท ลายสนิทเสรีกุล. คุณภาพของข้อสอบปรนัย ภาคเรียนที่ 1 ปีการศึกษา 2536 คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. จุฬาลงกรณ์เวชสาร 2537 กรกฎาคม;38(7): 391-398

การวิจัยเชิงพรรณนานี้ มีวัตถุประสงค์เพื่อศึกษาคุณภาพของข้อสอบปรนัย ของคณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ใช้ในภาคเรียนที่ 1 ปีการศึกษา 2536 ในรูประดับความยากง่ายและอำนาจจำแนกของข้อสอบ ความเที่ยงของข้อสอบ ตลอดจนจำแนกจำนวนของข้อสอบที่ดีเยี่ยม ข้อสอบที่ดี ข้อสอบที่ควรได้รับการแก้ไข ข้อสอบที่ไม่ดี และข้อสอบเชิงลบ ตามเกณฑ์มาตรฐานขององค์การอนามัยโลก วัสดุที่ใช้ในการศึกษาเป็นแบบสอบจำนวน 32 ฉบับ นำมาวิเคราะห์ข้อสอบด้วยโปรแกรม CTIA พบว่า จากแบบสอบ 32 ฉบับ เป็นข้อสอบปรนัยรวม 2,866 ข้อ ข้อสอบจำนวน 2,703 ข้อ (ร้อยละ 94.31) เป็นข้อสอบปรนัยชนิดมีคำตอบที่ถูกที่สุดเพียงคำตอบเดียว ข้อสอบจำนวน 163 ข้อ (ร้อยละ 5.69) เป็นข้อสอบปรนัยชนิดมีตัวเลือกถูกมากกว่า 1 ตัวเลือก ความเที่ยงของแบบสอบมีค่าระหว่าง 0.132-0.914 แบบสอบจำนวน 19 ฉบับ (ร้อยละ 59.38) มีความเที่ยงเท่ากับหรือมากกว่า 0.70 แบบสอบจำนวน 13 ฉบับ (ร้อยละ 40.63) มีความเที่ยงน้อยกว่า 0.70 ในจำนวนข้อสอบปรนัย 2,866 ข้อ เป็นข้อสอบที่ดีเยี่ยมจำนวน 480 ข้อ (ร้อยละ 16.75) เป็นข้อสอบที่ดีจำนวน 219 ข้อ (ร้อยละ 7.64) เป็นข้อสอบที่ควรได้รับการแก้ไขจำนวน 525 ข้อ (ร้อยละ 18.32) เป็นข้อสอบที่ไม่ดีจำนวน 1,403 ข้อ (ร้อยละ 48.95) และเป็นข้อสอบเชิงลบจำนวน 239 ข้อ (ร้อยละ 8.34) ข้อสอบปรนัยที่ดีเยี่ยมและที่ดีสมควรที่จะนำมาจัดเก็บเป็นคลังข้อสอบต่อไป

The multiple-choice question (MCQ) is the most flexible of the objective test types. It can be used to appraise the achievement of any of the educational objectives that can be measured by a paper-and-pencil test except those relating to skill in written expression and originality. An ingenious and talented writer of test questions can construct multiple-choice questions that require not only the recall of knowledge but also the use of skills of comprehension, interpretation, application, analysis, or synthesis to arrive at the keyed answer.[1] Multiple-choice examinations are being used more and more frequently in various fields of medicine for many purposes. They are used in medical schools as departmental examinations or as comprehensive examinations at the end of a school year; they are also used by several specialty boards as part of the testing procedure to determine certification for specialty practice.[2] While many formats of MCQ have been described, two types have been more widely used than others. Five-choice Completion (Type a), the simplest and most widely used multiple-choice form, consists of a question or incomplete statement followed by several (usually four or five) suggested answers or completions. Examinees are directed to select the one best answer.[3] In the second type, Multiple Completion (Type k), a common stem is followed by four statments, with one or more of them being correct. For example,the examinee is directed to select the A respones if 1, 2 and 3 are correct; B if 1 and 3 are correct; C if 2 and 4 are correct; D if only 4 is correct; and E if all four are correct.[4]

Good multiple-choice questions take time and skill to construct. Item analysis technique constitutes one of the most valuable tools that a classroom teacher can apply in attempting to improve the quality of his or her tests.[5] Every question is analyzed individually. This study analyze is the records of how many students chose the correct answer, how many chose the alternatives, and how many omitted to choose an answer at all. Although some alternative answers or "distractors" are ineffectual, some clearly attract many students. Examiners have a rich source of analytical data from which to review an examination in order to ensure that its fairness and accuracy are as impeccable as their science in their own discipline.[6] In the Faculty of Medicine, Chulalongkorn University, only one paper studied quality of the MCQ test used in the academic year 1979 are a comprehensive examination.[7] In view of this fact, the author was interested in studying the quality of MCQ tests used in first semester of academic year 1993.

Objectives

1. To study the quality of MCQ tests used in the first semester of academic year 1993 in terms of item-difficulty index, item-discrimination index and reliability of the tests.

2. To identify and count the number of excellent, good, revise, poor and negative questions by using standard criteria established by the World Health Organization (WHO).[8]

WHO definitions

1. **Excellent question** is one which has a discrimination index in the range of 0.35 and over, and a difficulty index between 0.30 and 0.70.

2. **Good question** is one which has a discrimination index in the range of 0.25 to 0.34 and a difficulty index between 0.30 and 0.70.

3. **Revise question** is one which has a discrimination index in the range of 0.25 and over and a difficulty index is not lying between 0.30 and 0.70.

4. **Poor question** is one which has a discrimination index under 0.25.

5. **Negative question** is one which has a negative discrimination index. Such a question is marked correct more frequently by bad than good students.

Materials

1. One microcomputer set (386 32-bit).
2. One printer set (EPSON LX-86).
3. One optical reader set (OPSCAN Model 5).
4. Sotfware packages :
   4.1 TOOLS : Software for the optical reader machine.
   4.2 CTIA : Software for item analysis.
   4.3 QEdit : Software for word processing.
5. Computer MCQ answer sheets.

Methods

1. Informed Faculty Committee and 20 department heads, Faculty of Medicine, Chulalongkorn University, about item analysis service before the time of the final examination of the first semester in academic year 1993.

2. Using optical reader machine, scanned students answer sheets in order to obtain raw data.

3. Using word processing equipment, prepared raw data which fit the CTIA format.

4. Running the item analysis program, identified questions in each MCQ test.

5. Counted the number of excellent, good, revise, poor, negative questions and calculated each of there categories as a percentage of the whole.

Results

1. There were 32 MCQ tests to be analysed by the item analysis program. The total number of questions amounted to 2,866 questions, of which 2,703 questions (94.31%) were the one best answer type and 163 questions (5.69%) the k type. twenty-four tests were used in the pre-clinical year, four tests in the clinical year, one test in the residency training program and three tests in the master of sciences program. The reliability of the 32 tests was in the range of 0.132-0.914. Nineteen tests (59.38%) had a reliability level of 0.70 and over. Details of the tests are shown in tables 1 and 2.

Table 1. Number of MCQ tests, items and percentage for each educational level.

Level	No. of tests	No. of question	Percentage
Pre-clinical	24	2,287	79.81
Clinical	4	279	9.73
Resident	1	150	5.23
M.Sc.	3	150	5.23
Total	**32**	**2,866**	**100.00**

Table 2. Number of questions and reliability of 32 MCQ tests.

Subject No.	Number of questions	Reliability
Pre-clinical	2,287	
1	100	0.515
2	100	0.888
3	82	0.841
4	99	0.874
5	100	0.864
6	100	0.686
7	100	0.686
8	100	0.725
9	120 (One 77, k 43)	0.847
10	84	0.816
11	130	0.847
12	50	0.732
13	100	0.850
14	107	0.865
15	100	0.648
16	120	0.867
17	112	0.914
18	112	0.690
19	50	0.508
20	66	0.841
21	60	0.282
22	120	0.795
23	75 (One 48, k 27)	0.706
24	100	0.815

Clinical	279	
1	29	0.132
2	120	0.650
3	30	0.164
4	100	0.719
Resident	150	
1	150 (One 57, k 93)	0.813
M.Sc.	150	
1	50	0.401
2	50	0.224
3	50	0.242

2. In using WHO's criteria to identify the quality of the 2,866 MCQ questions, 480 questions (16.75%) were judged an excellent, 219 (7.64%) good , 525 (18.32%) revise, 1,403 (48.95%) poor, and 239 (8.34%) negative [Table 3]. In comparing between the MCQ a type and the MCQ k type, the a type, or one best answer type, contained 447 excellent questions (16.54%), 205 good questions (7.58%), 502 revise questions (18.57%), 1,336 poor questions (49.43%) and 213 negative questions (7.88%). The MCQ k type contained 33 excellent questions (20.25%), 14 good questions (8.59%), 23 revise questions (14.11%), 65 poor questions (41.10%) and 26 negative questions (15.95%) [Table 4]. In assessing the tests by educational level, the pre-clinical MCQ tests had 380 excellent questions (16.62%), 170 good questions (7.43%), 431 revise questions (18.85%), 1,150 poor questions (50.28%) and 156 negative questions (6.82%). The clinical MCQ tests had 51 excellent questions (18.28%), 26 good questions (9.32%), 39revise questions (13.98%), 129 poor questions (46.24%) and 34 negative questions (12.18%). The residency MCQ tests had 36 excellent questions (24.00%), 7 good questions (4.67%), 23 revise questions (15.33%), 56 poor questions (37.33%) and 28 negative questions (18.67%). The M.Sc. MCQ tests had 13 excellent questions (8.67%), 16 good questions (10.67%), 32 revise questions (21.33%), 68 poor questions (45.33%) and 21 negative questions (14.00%) [Table 5].

Table 3. Quality of the MCQ test questions.

Quality of question	Number of question	Percentage
Excellent	480	16.75
Good	219	7.64
Revise	525	18.32
Poor	1,403	48.95
Negative	239	8.34
Total	**2,866**	**100.00**

Table 4. Quality of questions in one best type and k type tests.

Type	Number of question	Percentage
One best answer	2,703	100.00
Excellent	447	16.54
Good	205	7.58
Revise	502	18.57
Poor	1,336	49.43
Negative	213	7.88
k type	163	100.00
Excellent	33	20.25
Good	14	8.59
Revise	23	14.11
Poor	67	41.10
Negative	26	15.95

Table 5. Quality of MCQ tests used in Pre-clinical, Clinical, Residency Training and Master of Sciences program.

Quality of question	Number of questions	Percentage
Pre-clinical	2,287	100.00
Excellent	380	16.62
Good	170	7.43
Revise	431	18.85
Poor	1,150	50.28
Negative	156	6.82
Clinical	279	100.00
Excellent	51	18.28
Good	26	9.32
Revise	39	13.98
Poor	129	46.24
Negative	34	12.18
Resident	150	100.00
Excellent	36	24.00
Good	7	4.67
Revise	23	15.33
Poor	56	37.33
Negative	28	18.67
M.Sc.	150	100.00
Excellent	13	8.67
Good	16	10.67
Revise	32	21.33
Poor	68	45.33
Negative	21	14.00

Discussion

Of the total of 2,866 question contained in 32 MCQ tests, 699 (24.39%) were judged as containing questions of high quality (excellent questions and good questions). It is most useful in preparing such questions to have access to a question bank. Revise questions are questions with high discrimination indexes. They can help teachers to determine who are the best students; however, they pose a problem in that their difficulty index is not between 0.30 and 0.70. This criterion is based on a WHO suggestion,[8] but in general, a question with a difficulty index between 0.20 and 0.80 is acceptable.[9,10] A good question much discriminate between examinees who are more competent and those who are less competent with respect to the particular point in a question. In our study there were 1,403 poor questions (48.95%), i.e. questions that were either too easy or too difficult. If a question is too easy,[11] so that it is answered correctly by all examinees, it fails to discriminate between good students and poor students and adds nothing to the test. If thequestion is too difficult, the proportion of students answering it correctly may be close to the proportion of those who might be expected to select the right answer by chance alone. Other reason for poor quality is distractors, which should be equally plausible to the naive candidate.[12] If not, they would be disregarded and the question effectively becomes one from 2, 3, or 4 rather than from 5. A good way to derive distractors is first to think of the common errors students make and the common misconceptions they hold. The negative question must be a crucial point about which all teachers should be concerned. A question scored correct more frequently by bad than good students should be examined carefully to clarify why good students are not choosing the correct answer.[6] Perhaps the question is out of date, perhaps it can be interpreted in more than one way, or perhaps the teaching differs concerning what is correct. The best way to improve the revise, poor and negative questions is the establishment of working committees made up of subject-matter specialists who meet with the test specialists like the comprehensive examination committee. The member who be the department representative can inform and feedback to the department and each instructor. This is the first step in question improvement.[13,14]

All certifying examinations should have a reliability level above 0.70.[15-17] Examinations with reliability below 0.60 are unfair. Reliability of an examination increases with the number of question in the test, the homogeneity of the questions, the number of questions of high discrimination and of average difficulty, and the range of ability among candidates.[18] Although about 60 percent of the MCQ tests had a reliability level above 0.70, 40 percent of them, i.e. those with a reliability below 0.70, should be reconsidered by examiners. If the reliability of the test is too low, it serves as a signal that something is wrong with the test : it may not contain enough items, or, to be more precise, it may not contain enough effective questions [19] A way to increase the reliability of the test is writing more representative test question. The longer the test, (the test has many questions) the greater its reliability.[20] The Faculty should have policy on training for all instructors about how to construct a fair multiple choice question.

Currently, machine analysis can provide quite specific feedback on the class and the individual, on the examination as a whole and on each question. Teachers have no excuse for not providing feedback to students on their performance.[6]

Summary

This descriptive research studied the quality of MCQ test used during the first semester of academic year 1993, at the Faculty of Medicine, Chulalongkorn University in terms of item-difficulty index, item-discrimination index, reliability of tests and identification of the number of excellent, good, revise, poor and negative questions they contained by using standard criteria of World Health Organization. Thirty-two MCQ tests were analysed by the CTIA item analysis program. There were 2,866 items : 2,703 (94.31%) of them were the "one best answer" type and 163 items (5.69%) were the k type. The reliability of the tests was in the range of 0.132-0.914. Nineteen tests (59.38%) had a reliablity equal to 0.70 and over and 13 tests (40.63%) had a reliability below 0.70. Among 2,866 questions, there were 480 excellent questions (16.75%), 219 good questions (7.64%), 525 revise questions (18.32%), 1,403 poor questions (48.95%) and 239 negative questions (8.34%). Both excellent and good questions are most useful in the preparation of a question bank.

References

1. Thorndike RL, Hagen EP. Measurement and Evaluation in Psychology and Education. 4th.ed. New York: John Wiley & Sons, 1977.228
2 Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine :A Guide for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.5
3. Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine : A Guide for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.15

4. Hubbard JP, Clemans WV.Multiple-Choice Examination in Medicine: A Guide for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.27
5. Payne DA. The Specification and Measurement of Learning Outcomes. Toronto: Xerox College Publishing,1968.145
6. Cox K,Ewan CE. The Medical Teacher. 2nd.ed. London: Churchill Livingstone, 1988.164
7. Phulklongtan M, Jaroongdaechakul M, Limpapayom K. An Analysis of Comprehensive Examination Items of Academic Year 1979. Chula Med J 1981 Sep; 25(5):1035-40
8. Guilbert JJ. Educational Handbook for Health Personnel. Revised ed. Geneva: World Health Organization, 1987.4.69
9. Kasemsunti C, Wachirapornthip A. Item analysis. Siriraj Hosp Gaz 1975 Aug; 27(8):1181-96
10. Sukamolson S. Classical Test Item Analysis and Grading Manual (V.6.30). Bangkok: Chulalongkorn University Language Institute, 1992.36
11. Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine : A Gulde for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.39
12. Cox K, Ewan CE. The Medical Teacher. 2nd.ed. London: Churchill Livingstone, 1988.155
13. Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine: A Guide for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.31
14. Cox K, Ewan CE. The Medical Teacher. 2nd.ed. London: Churchill Livingstone, 1988.166
15. Schumacher CE. Scoring and analysis. In : Hubbard JP,ed. Measuring Medical Education. Philadelphia: Lea & Febiger, 1971.
16. Cox K, Ewan CE. The Medical Teacher. 2nd.ed. London : Churchill Livingstone, 1988.163
17. Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine : A Guide for Examiner and Examinee. Philadelphia : Lea & Febiger, 1961.71
18. Ebel RL. Essentials of Educational Measurement. New Jersey : Prentice Hall, 1972.
19. Hubbard JP, Clemans WV. Multiple-Choice Examinations in Medicine: A Guide for Examiner and Examinee. Philadelphia: Lea & Febiger, 1961.72
20. Ary D, Jacobs LC, Razavieh A. Introduction to Research in Education. 2nd.ed. New York: Holt, Rinehart and Winston, 1979.215