Original Research Article

True-false analysis reveals inherent flaws in multiple true-false tests

Thomas Puthiaparampil[1]*, Henry Rantai Gudum[2], M. Mizanur Rahman[3], Rosalia Saimon[3], Isabel Fong Lim[4]

[1]Department of Medicine, [2]Department of Pathology, [3]Department of Community Medicine and Public Health, [4]Department of Paraclinical Sciences, Faculty of Medicine and Health Sciences, UNIMAS, Sarawak, Malaysia

*Correspondence:
Dr. Thomas Puthiaparampil,
E-mail: pthomas@unimas.my

ABSTRACT

Background: Multiple true-false tests (MTF), a component of our assessment system, have consistently generated low scores and many failures. This was attributed to the negative marking scheme. However, no study was conducted to explore the issue further. Item analysis revealed that students omitted false options more frequently and answered them wrongly more frequently than true options. The aim of this study was to determine the performance discrepancy between true and false options of MTF tests and the reasons for such discrepancy and the poor performance of MTF in general.
Methods: The student performance of past 7 years of year-3 medicine end-of-posting examinations was analysed. The item analysis reports of 23 MTF tests were used to determine the significance of the differences in omission rates, correct-answer rates and the discrimination index of true and false options.
Results: There were statistically significant differences in the omission rates, correct-answer rates and discrimination index values of true and false options. This study revealed that the false options consistently let down student performance. Although negative marking could be partly blamed for the situation, no justification could be found for the use of false options to test knowledge.
Conclusions: Some publications endorse MTF, but many highlight its drawbacks. The use of false options in MTF was seen as an inherent defect in this instrument. As viable alternatives like VSAQ and Constructed Response Tests are in the horizon, we conclude that MTF ought to be discarded as an assessment instrument.

Keywords: Multiple true-false tests, True-false analysis, Flaws in MTF tests

INTRODUCTION

It is observed that many institutions have discarded multiple-choice true-false (MTF) tests in favour of single best answer questions (BAQ) and other types of assessments.[1] Our faculty also has decided to phase out MTF. Our MTF items consist of a stem and five options, of which zero to any number could be true or false. The student is required to answer them as true, false or don't know (omit). One mark is awarded to a correct answer and one negative mark to an incorrect answer, while the omitted items get zero mark. The maximum mark awarded for one question is 5 and the minimum zero. Usually 20 to 60 questions are used in a test in our faculty. The optical mark reader (OMR) by SmartScan, which also generates item analysis reports, does the scoring. This report include the discrimination index (DISi) of each option as well as that of each question, how many candidates omitted each option, and the percentage of candidates who got each option correct [difficulty index (DIFi)]. MTF tests have been used in our assessment system ever since the inception of the Faculty

of Medicine and Health Sciences, Universiti Malaysia Sarawak (UNIMAS) over 2 decades ago. However, it has not been used as the sole assessment, either formative or summative. Other assessment instruments used are One Best answer questions (BAQ), Modified essay questions (MEQ), Short essay questions (SEQ), Short answer questions (SAQ), Objective structured clinical examinations (OSCE), and clinical assessments of long case and short cases using real patients.

MTF tests suffer from two major drawbacks, guessing and the cueing effect. Guessing could be discouraged by negative marking, however, guessing ability is unrelated to the subject being tested.[2] Cueing can be difficult to disentangle from guessing but has been estimated to play a role in approximately 20% of answers.[3] Swanson et al found that, after reviewing 'literally tens of thousands' of true-false MCQs, they were difficult to write well.[4] In addition, in order to avoid ambiguity, the writer is pushed to assessing the recall of an isolated fact, thus, unfair to an otherwise competent student who may fail if he-she has not memorized isolated facts. It has been noticed that our students consistently scored poorly in MTF compared to other components of the assessments. Other authors also have made similar observations.[5] This issue was considered inherent to MTF, and the poor scores were attributed to negative marking used as a deterrent to guessing. However, no study was conducted in our setting to explore the issue further. Considering this context, this study was aimed to determine the performance discrepancy between true and false options of MTF tests used in our faculty, and to solicit reasons for such discrepancy and the poor performance of MTF in general.

METHODS

The project was conceptualized and proposal submitted for faculty ethics committee approval in February 2016. The study proposal was approved by the Technical Review Committee (TRC) of the Faculty of Medicine and Health Sciences (FMHS), Universiti Malaysia Sarawak (UNIMAS). Ethical clearance was also obtained from the Institutional Review Board (IRB) of the Faculty. Strict anonymity and confidentiality of the data were maintained. No student particulars were extracted and disclosed publicly. The data collection and the study process were delayed for two years and finally the data analysis and writing up got completed in July 2019. The study material included all available posting and professional examination results of undergraduate medical students irrespective of gender and year of study from 2012 to 2018.

The student scores in MTF, BAQ, MEQ and OSCE tests of year-3 medicine end of posting examinations of the academic years 2012-13, 2013-14, 2014-15, 2015-16, 2016-17, 2017-18 and 2018-19 (a total of 23 examinations) were statistically analysed to determine the pattern in student performance. Other materials used were

the item analysis reports of 23 MTF tests, which were used in the year 2 first professional examination, year 3 (end of posting examinations of medicine and paediatrics) and year 5 (paediatrics end of posting examination and final professional examinations) of our faculty's medical programme. The total number of MTF items studied was 550 with a total of 2750 options. Among them 58% were true options and 42% false options.

We performed independent-sample t tests on SPSS version 22 and determined the p values of the following 4 performance parameters of true and false options of the MTF tests: proportion of options omitted, proportion of options answered correctly, proportion of options answered incorrectly and means of discrimination index of true options and false options in each of the tests. A p value of <0.05 was considered to be statistically significant.

RESULTS

Students' mean scores in four assessments in year 3 medicine posting examinations of 7 years showed MTF to be the lowest with a decline in its scores over the years. Although there was a declining trend, the other 3 assessments remained consistently above MTF (Figures 1 and 2).
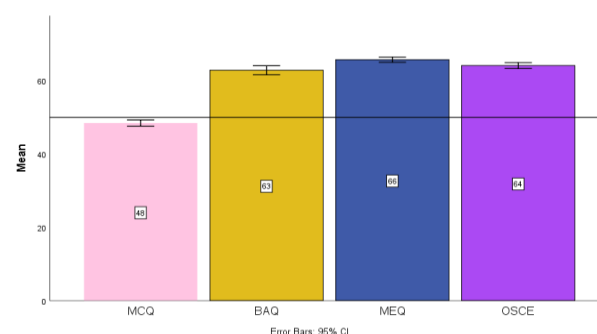

Figure 1: Mean distribution of student scores in 23 examinations of year-3 medicine.
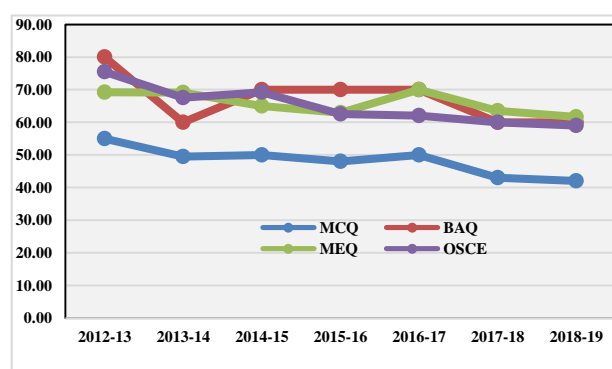

Figure 2: Median of student scores in 23 examinations of year-3 medicine over 7 years.

The proportion of omission was higher in false options compared to true options in 21 out of 23 tests, out of which 8 showed statistical significance. The proportion of options answered correctly was higher in true options compared to false options in 22 out of 23 tests, out of which 21showed statistical significance. The proportion of options answered incorrectly was higher in false options compared to true options in 22 out of 23 tests, out of which 20 showed statistical significance. The means of discrimination index values were higher in false options compared to true options in 22 out of 23 tests, out of which 14 showed statistical significance (Table 1).

Table 1: The data of 23 MTF tests.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N	%	%	%	%		%	%		%	%		Mean	Mean	
100	59	41	18.22	21.73	0.29	71.02	48.22	0	10.64	30.12	0	0.11	0.21	0.03
100	62	38	19.68	25.29	0.12	73.71	52.08	0	6.55	22.61	0	0.08	0.17	0.02
100	58	42	16.19	23.71	0.02	75.05	56.14	0	8.69	19.93	0	0.15	0.26	0.01
100	51	49	16.84	22.53	0.08	74.04	59.94	0.00	9.06	17.53	0	0.07	0.19	0.00
100	57	43	20.14	20.84	0.84	67.33	50.91	0.00	12.42	28.07	0	0.13	0.19	0.09
100	62	38	19.29	20.21	0.81	72.87	69.37	0.52	7.73	10.24	0.37	0.15	0.15	0.87
100	56	44	22.54	26.66	0.23	69.34	51.02	0	8.18	22.25	0	0.14	0.32	0
100	53	47	16.62	26.62	0.00	74.55	46.11	0	8.92	27.36	0	0.08	0.21	0.00
100	64	36	0.33	0.22	0.53	74.3	81.06	0.05	25.35	18.72	0.05	0.16	0.16	0.86
100	53	47	24.68	31.72	0.08	67.57	45.49	0	7.91	23.04	0	0.12	0.17	0.18
150	55	45	21.16	29.22	0.03	70.98	49.44	0	7.82	21.44	0	0.17	0.32	0
100	60	40	19.92	26.88	0.04	68.55	51.4	0	11.48	21.8	0	0.16	0.20	0.43
100	64	36	20.19	24.67	0.17	70.98	55.31	0.00	8.8	19.94	0	0.11	0.32	0
100	53	47	16.62	26.62	0.00	74.55	46.11	0	8.92	27.36	0	0.08	0.21	0.00
100	57	43	14.07	20.63	0.02	75.28	52.67	0	10.6	26.6	0	0.06	0.22	0
100	62	38	17.42	16.82	0.83	71.94	59.87	0.01	10.63	23.34	0	0.15	0.19	0.35
100	55	45	18.82	22.07	0.25	68.6	47.38	0	12.49	30.49	0	0.11	0.15	0.30
100	54	46	16.89	20.28	0.25	73.2	57.74	0	9.91	22.04	0	0.10	0.28	0
100	64	36	17.17	23.06	0.09	71.14	57.31	0.01	11.73	19.64	0.05	0.16	0.2	0.27
100	54	46	17.48	21.72	0.1	70.98	47.74	0	11.46	30.52	0	0.21	0.32	0.01
300	60	40	24.83	31.24	0.00	64.87	38.34	0	10.44	30.47	0	0.07	0.14	0.01
300	55	45	24.75	30.13	0.01	64.69	44.65	0	10.57	25.26	0	0.12	0.18	0
100	62	38	22.71	27.16	0.27	71.05	46.13	0	6.06	26.47	0	0.15	0.24	0.05

Specifics of the 15 columns in this table are: 1. Number of options in the test (number of questions x 5); 2. Percentage of true options in the test (T); 3. Percentage of false options in the test (F); 4. Mean of the percentage of students omitting true options in the test. e.g., if 12 out of 29 students omitted an option=(12/29)×100=41.38%. Then the mean of all the omission percentages of true options in the test is calculated (TO); 5. Similar calculation for false options in the test (FO); 6. P value of the column 4 and 5; 7. Mean of the percentage of true options answered correctly in the test (TC); 8. Mean of the percentage of false options answered correctly in the test (FC); 9. P value of the 7 and 8; 10. Mean of the percentage of true options answered wrongly in the test (TW). 11. Mean of the percentage of false options answered wrongly in the test (FW); 12. P value of the above 10 and 11; 13. Mean discrimination index of all true options in the test (DISi T); 14. Mean discrimination index of all false options in the test (DISi F); 15. P value of the above 13 and 14. (T=true options, F=false options, TO=true options omitted, FO=false options omitted, TC=true options answered correctly, FC=false options answered correctly, TW=true options answered wrongly, FW=false options answered wrongly, DISi=discrimination index, n=number, p <0.05, omission 25%, TC and FC 70%, TW, FW 25%, DISi 0.2 are significant values).

DISCUSSION

Consistently low student scores of MTF tests in our examinations have raised concerns. The literature is inconsistent in this regard, as many authors acclaim MTF, while others have moved away from it to other assessment options like One Best MCQ and open-ended very short answer questions. The proponents of MTF argue that the problem is not in the instrument, but in the way it is used.[6,7] We would agree with the argument, that if improperly used, any instrument would perform badly. But our past experience does not offer hope of MTF recuperating with time from its current dismal status. The perceived advantages of MTF are: easier than BAQ to construct, which is debatable; the five options of an MTF could test five independent items or facts in contrast to BAQ, which tests only one fact per item. Thus, the wide coverage of topics possible in MTF claims to improve the reliability and validity of the test.[8-11]

Case and Swanson (2001) explored why MTF format has been abandoned in the USA.[4] After reviewing 'literally tens of thousands' of true or false MCQs, they found that they are not only difficult to write well but, in order to avoid ambiguity, the writer is pushed to assessing the recall of an isolated fact. An otherwise competent student

may fail if he or she has not memorized isolated facts, thus, such a format is therefore unfair.[1] Abortive attempts were made to test skills, attributes and areas of learning that cannot properly be assessed by MCQs (MTF), and such inappropriate use is one of the main reasons why MCQs have been attacked so often.[6] As a result of this misguided enthusiasm, far too many bad questions were written in the early days with ungrammatical and ambiguous wording, lack of precision, and a demand for knowledge that was on the one hand trivial, and on the other hand far too detailed and often irrelevant.

Writing crystal clear items is a real challenge.[9] Overuse and inappropriate use of MTF has spoiled its reputation. Good candidates were more affected by the ambiguities in the question. Personal opinions of experts may become a problem in the questions.[6] Our results showed that the false options in MTF tests performed worse in statistically significant proportions than the true options in all parameters except in discrimination index. The false option discriminating good students from poor students better than true options is understandable. Among the average students, the false options were less frequently answered correctly than true options and were omitted more frequently than true options. Students would omit an option because either they do not understand it, or they are unsure of it being true or false. As each MTF test contains nearly 50% false options, the poor performance in false options lowers students' scores considerably. MTF being one of the components in our decisive examinations, students' final grades were also adversely affected by it.

There are several reasons why false options perform worse than true ones. True statements are what students read in textbooks and what they are taught in teaching sessions. False options are neither found in textbooks nor taught by the lecturers, thus, when the false statements suddenly appear in the question papers, the students are unsure. Whereas true statements are instantly decidable, it takes much more thinking, searching and sifting through existing knowledge to be confidently sure that a false statement is really false. The students omit the options, if there is an iota of doubt because of the scare of negative marking. It is argued that discarding the intimidating effect of negative marking would solve the problem, and it seems the negative marking of MTF is on the way out.[8,12] If negative scoring is discarded, a student who blindly answers all options 'as true' might get at least 50% marks, as more than 50% options are likely to be true in a paper. The passing mark will need to be raised from 50 to 75% or standard setting applied. Moreover, the discriminating power of MTF would diminish, as the test would become easier.

Another pertinent issue is the quality of items. Are false options more difficult than true options to construct? The answer is more likely to be yes, as they have no basis in textbooks. They are up to the whims and fancies of the question authors. It appears that many times the false options are written with latitude, as if they do not matter, as they are false anyway. It may appear of no consequence to the question author, while it makes a huge difference to the student who is given the task of confidently judging the option as false. Sometimes the false options are outright false, which becomes a giveaway with no discriminating power. Since the false options are rather artificially generated by question authors, they are often flawed. Sometimes they are out rightly absurd, when the writer just turns around a true statement into a false one. Test-wary students and the 'deep thinking' students end up in problem, suspecting whether there is something tricky there. Confusion can arise from every word the question author uses carelessly, no matter true or false the option. The amount of thinking, care and scrutiny needed in question writing cannot be overemphasized.

MTF can be perceived as a compromise instrument with the sole advantage of machine marking. While MEQ, SEQ and SAQ demands the student to write the answers, MTF requires them to write nothing but judge the options given. While the former instruments bring out the knowledge and thought process of the students, which are helpful in giving feedback and modifying the teaching/learning process, MTF gives no such information. Open-ended assessments like very short answer questions (VSAQs) and Constructed Response Examinations are upcoming methods which require the students to write the answers.[13,14] They have the disadvantage of no machine marking but offers advantages like knowing students' depth of knowledge and grasp of the subject, which are useful for providing feedback and modifying the teaching/learning process.

CONCLUSION

This study concludes that, while there is truth in the arguments in favour, the arguments against MTF are overriding. Although this instrument was extensively used in the past, there appears to be inherent defects in the instrument, which appear difficult to overcome even with extensive experience and expertise in its use. Moreover, MTF fails to bring out the thought process and depth of knowledge of the test takers, which are essential components to help improve the teaching/learning process. With the availability of very short answer questions and constructed response tests as viable alternatives. We recommend a closure of MTF tests. Data used in this study belong to a single medical faculty. We did not find any other study comparing the performance discrepancy between true and false options. This may be considered a limitation of the study.

ACKNOWLEDGEMENTS

examination results and item analysis reports for this project.

REFERENCES

1. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 2nd edition. Philadelphia: National Board of Medical Examiners; 1998: 22-25.
2. Jolly B. Assessment and examination. Adv Psychiatric Treat. 1999;5:405–14.
3. Schuwirth LWT, Van Der Vleuten CPM, Donkers HHLM. A closer look at cueing effects in multiple-choice questions, Med. Edu. 1996;30:44–9.
4. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 3rd edition. Philadelphia: National Board of Medical Examiners; 2001.
5. Rasiah SMS, Isaiah R. Relationship between item difficulty and discrimination indices in true or false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore. 2006;35(2):67-71.
6. John A. For Multiple Choice Questions Medical Teacher. 1979;1:37-42.
7. John A. The MCQ Controversy: a Review Medical Teacher. 1981;3:150-6.
8. John A. Medical Teacher 25th anniversary series multiple choice questions revisited. Medical Teacher. 2004;26(2):110-3.
9. Leonard AB. Hints for students (and examiners) on answering MCQ questions of the multiple true-false type, Medical Teacher.1986;8(1):41-8.
10. Smart GA. The multiple choice examination paper. British J Hospital Med. 1976;15:131.
11. Tavakol M, Dennick R. Post-examination analysis of objective tests, Medical Teacher. 2011;33:447-58.
12. Premadasa IG. A reappraisal of the use of multiple choice questions. Medical teacher. 1993;15(2-3):237-42.
13. Hauer KE, Boscardin C, Brenner JM, van Schaik SM, Papp KK. Twelve tips for assessing medical knowledge with open-ended questions: Designing constructed response examinations in medical education. Medical teacher. 2019;5:1-6.
14. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. BMC Med Edu. 2016;16(1):266.