

Multiple-Choice Questions in Preclinical and Clinical Phase Examination for Undergraduate Medical Program in a Malaysian Medical School: A 2-Year Analysis

Siti Khadijah Adam (✉ sk.adam@upm.edu.my)

University Putra Malaysia <https://orcid.org/0000-0003-2716-5705>

Faridah Idris

Universiti Putra Malaysia

Puteri Shanaz Jahn Kassim

Universiti Putra Malaysia

Nor Fadhlina Zakaria

Universiti Putra Malaysia

Rafidah Hod

Universiti Putra Malaysia

Research article

Keywords: Item analysis, Multiple choice question, Difficulty index, Discrimination index, Distractor efficiency

Posted Date: July 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38264/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multiple-choice questions in preclinical and clinical phase examination for undergraduate medical program in a Malaysian medical school: A 2-year analysis

Siti Khadijah Adam^{1,2*}, Faridah Idris^{1,3}, Puteri Shanaz Jahn Kassim^{1,4}, Nor Fadhlin Zakaria⁵, Rafidah Hod^{1,2}

¹Medical Education Research and Innovation Unit, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.

²Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.

³Department of Pathology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.

⁴Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.

⁵Department of Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia.

*** Corresponding author: Siti Khadijah Adam**

Medical Education Research and Innovation Unit, and

Department of Human Anatomy,

Faculty of Medicine and Health Sciences,

Universiti Putra Malaysia,

43400 Serdang,

Selangor Darul Ehsan,

Malaysia.

Tel: +603-97692354

Fax: +603-97692341

E-mail: sk.adam@upm.edu.my

Abstract

Background: Multiple-choice questions (MCQs) are used in measuring the student's progress and post-examination analysis is usually done to guarantee the item's appropriateness for question banking. Item analysis is a simple and effective method to determine the quality of MCQs by using three parameters; difficulty or passing index (PI), discrimination index (DI) and distractor efficiency (DE).

Methods: This study analysed the MCQs in the preclinical and clinical examinations in Doctor of Medicine program of Universiti Putra Malaysia. Forty MCQs consisted of four options each in the preclinical examination and 80 MCQs with five options from the clinical examination paper in 2017 and 2018 were analysed and compared.

Results: The mean DI was similar in all examinations, except a significant reduction in 2018 clinical examination. From 2017 and 2018, preclinical MCQs showed an increment in the number of 'excellent' and 'good' items. However, clinical papers showed reduction in DI due to high number of 'poor' questions. Comparing both years, there was an increase in the number of items with no non-functioning distractors in both examinations. Among all, preclinical MCQs in 2018 showed the highest mean of DE as compared to the others.

Conclusion: Our findings suggested that item authors from preclinical phase showed an improvement in constructing good quality MCQs while clinical phase authors may need more training and continuous feedback. Higher number of options did not affect the level of difficulty of a question however the discrimination power and effectiveness of distractors might differ.

Keywords: Item analysis, Multiple choice question, Difficulty index, Discrimination index, Distractor efficiency

1 **Background**

2 Doctor of Medicine (MD) curriculum in Universiti Putra Malaysia (UPM) consists of two phases;
3 pre-clinical and clinical phases which runs for 2 and 3 years, respectively. Students will sit for a
4 summative assessment at the end of each phase. Students must pass the preclinical examination
5 to proceed to clinical phase. In the fifth year, they need to sit and pass the clinical phase
6 examination before awarded with the degree of medicine. In the written examination, various
7 assessment tools are used such as MCQs, short answer questions and modified essay questions.

8 In any medical examinations, MCQ is one of the most important well-established written
9 assessment tools widely used for its distinct advantage and ability to evaluate over a broad
10 coverage of concepts in less time. The scoring is also objective and reliable [1]. The type A MCQ
11 item (single best response) consists of a 'stem' or 'vignette', followed by a 'lead in' statement and
12 several options. The correct answer in the list of options is called as a 'key' and the incorrect
13 options are called 'distractors'.

14 The validity of an assessment refers to the evidence presented to support or to refute the
15 meaning or interpretation assigned to the assessment data [2]. Validity, therefore, is a degree to
16 which the test measures what it is supposed to measure. This includes test item analysis which
17 is usually done after the assessment has completed to determine the candidate responses to
18 individual test items, the quality of those items as well as the overall assessment. Difficulty index
19 or passing index (PI), discrimination index (DI) and distractor efficiency (DE) of each item can be
20 obtained from the analyses which reflect the quality of the test items. The PI of an item is
21 commonly defined as the percentage of students who answered the item correctly. The DI, on the
22 other hand, is defined as the degree to which an item discriminates between students of high and
23 low achievement. The DE is used to assess the credibility of the distractors in an item, whether
24 they are able to distract students from selecting the right answer [3]. Any distractor that is selected
25 by less than 5% of the students is considered to be a non-functional distractor (NFD). The

1 advantages of the analysis are to help identify faulty items, identify the lower performers and their
2 learning problems such as misconceptions as a guide for remedial actions to be done to students
3 and as importantly to increase educator's skills to construct a high quality of test items.

4 For this reason, we analysed the MCQs given in the preclinical and clinical phase
5 examinations in the year of 2017 and 2018. A comparison was made between PI, DI and DE of
6 the items between the examinations for both years as well as between the two phases of the
7 examinations. Good quality items and revised items shall be stored in the question bank and faulty
8 items shall be discarded based on the obtained findings.

9 10 **Methods**

11 The study was conducted in the Faculty of Medicine and Health Sciences, UPM during the
12 preclinical and clinical phase examinations in the year 2017 and 2018. In 2017, a total of 84
13 second year MD students took the preclinical examination while 128 final year students took the
14 clinical phase examination. Meanwhile in 2018, there were 100 second year students and 120
15 final year students who took the end of preclinical and clinical examination, respectively. The 2-
16 year preclinical phase examinations comprised of 40 MCQs with four options each, while the 2-
17 year clinical phase examination comprised of 80 MCQs with five options each. Each correct
18 response was awarded five marks and there was no negative marking for the wrong answers.
19 Pre-validation of the items was done by the vetting committee of the faculty.

20 21 **Item analysis**

22 Post-validation was done automatically by item analysis using the OMR machine (Scantron
23 iNSIGHT 20 OMR scanner, Minnesota USA). The scores of all students in each examination

paper were arranged in order of merit. The upper 27% students were considered as 'top' students and lower 27% students as 'poor' students. Each item was analysed for difficulty and discrimination indices according to Hassan & Hod [4] and Abdul Rahim [5] :

(i) Difficulty index or Passing Index (PI), using the formula

$$PI = (H + L)/N$$

(ii) Discrimination index (DI), using the formula

$$DI = (H - L)/A$$

H = number of 'top' students answering the item correctly

L = number of 'poor' students answering the item correctly

N = total number of students in the 'top' and 'poor' groups

A = number of students in 27% of total students

The interpretation of PI and DI values are presented in Table 1.

(iii) Distractor efficiency (DE)

Non-functional distractor (NFD) is the option that was selected by less than 5% of students. Based on NFDs in an item, DE ranges from 0% to 100%. If an item with four options contains three or two or one or nil NFDs, then DE would be 0, 33.3%, 66.7% and

100.0%, respectively. If an item with five options contains four or three or two or one or nil NFDs, then DE would be 0, 25%, 50%, 75% and 100%, respectively.

Table 1 Interpretation of item analysis

Difficulty @ Passing Index (PI)	Values	Discrimination Index (DI)	Values
Difficult	≤ 0.3	Excellent	≥ 0.35
Ideal	0.31 – 0.59	Good	0.25 – 0.34
Acceptable	0.6 – 0.7	Acceptable	0.21 – 0.24
Very Easy	> 0.7	Poor	≤ 0.2

*Adapted from Ananthakrishnan [6]

Results

A total of 240 MCQs were analysed and the average PI and DI were determined. Overall, it was found that the difficulty level of the questions was similar in both preclinical and clinical phase examinations (Table 2). Interestingly, the values of average PI reduced in 2018, indicating a reduction of an increase of difficulty level of the questions for that particular year in both preclinical and clinical examinations. Nevertheless, the mean DI was similar in all examinations except for a low DI in the 2018 clinical phase examination.

For preclinical phase examination, the numbers of 'difficult' and 'very easy' items were similar in both years (Table 3). Half of the 40 MCQs were 'ideal' and 'acceptable'. Clinical phase examination in 2017 had 54% 'ideal' and 'acceptable' items but the value reduced in 2018 to 34% due to the significant increase in the percentage of 'difficult' items in the year 2018.

Table 2 Average of difficulty index and discrimination index of items in preclinical phase examination (n= 40) and clinical phase examination (n= 80) in 2017 and 2018

		Mean \pm SD			
		Preclinical Phase Examination		Clinical Phase Examination	
		2017	2018	2017	2018
Difficulty @ Passing		0.60 \pm 0.24	0.55 \pm 0.23	0.60 \pm 0.21	0.56 \pm 0.28
Index (PI)		(Acceptable)	(Ideal)	(Acceptable)	(Ideal)
Discrimination Index		0.25 \pm 0.16	0.31 \pm 0.16	0.25 \pm 0.15	0.20 \pm 0.11
(DI)		(Good)	(Good)	(Good)	(Poor)

Table 3 Summary of no of items in preclinical phase examination (n= 40) and clinical phase examination (n= 80) in 2017 and 2018 based on difficulty index (PI)

Difficulty @	Number of Items (%)			
Passing Index (PI)	Preclinical Phase Examination		Clinical Phase Examination	
	2017	2018	2017	2018
Difficult	7 (17.50)	7 (17.50)	7 (8.75)	21 (26.25)
Ideal	12 (30.00)	15 (37.50)	29 (36.25)	18 (22.50)
Acceptable	8 (20.00)	5 (12.50)	14 (17.50)	9 (11.25)
Very Easy	13 (32.50)	13 (32.50)	30 (37.50)	32 (40.00)

There was an increment in the total percentages of 'excellent' and 'good' items in preclinical phase examination in 2018, from about 38% to 70% (Table 4). In clinical phase examination of 2017, half of the questions were 'excellent' and 'good'. However, the percentage reduced to 36% in 2018 due to the high percentage of 'poor' questions in the examination (53%).

Additionally, there were five questions with zero DI in the paper; one with PI equals to 1 and another with PI equals to 0.

Table 4 Summary of number of items in preclinical phase examination (n= 40) and clinical phase examination (n= 80) in 2017 and 2018 based on discrimination index (DI)

Discrimination Index (DI)	Number of Items (%)			
	Preclinical Phase Examination		Clinical Phase Examination	
	2017	2018	2017	2018
Excellent	11 (27.50)	15 (37.50)	23 (28.75)	6 (7.50)
Good	4 (10.00)	13 (32.50)	17 (21.25)	23 (28.75)
Acceptable	9 (22.50)	3 (7.50)	10 (12.50)	9 (11.25)
Poor	16 (40.00)	9 (22.50)	30 (37.50)	42 (52.50)

The total number of non-functioning distractors (NFD) was reduced in 2018 in both examination phases (Table 5). Both examinations in 2018 showed an increase in the number of items with no NFD as compared to the previous year. The number of items with no NFD in preclinical phase examination is higher than clinical phase examination for both years. Similarly, the overall mean DE was increased in 2018 with preclinical phase examination showing the highest mean DE as compared to the rest of examinations.

Table 5 Distractor efficiency (DE) of multiple-choice questions in preclinical phase examination and clinical phase examination in 2017 and 2018

Parameter	Number of Items (%)			
	Preclinical Phase		Clinical Phase	
	Examination		Examination	
	2017	2018	2017	2018
Number of items	40		80	
Distractors	120		320	
Functioning distractors	83 (56.67)	86 (71.67)	155 (48.44)	177 (55.31)
Non-functioning distractors	52 (43.33)	34 (28.33)	165 (51.56)	143 (44.69)
Items with no NFD (DE = 100%)	9 (22.50)	13 (32.50)	3 (3.75)	8 (10.00)
Items with 1 NFD (DE = 66.67%)	15 (37.50)	20 (50.00)	21 (26.25)	24 (30.00)
Items with 2 NFDs (DE = 33.33%)	11 (27.50)	7 (17.50)	28 (35.00)	32 (40.00)
Items with 3 NFDs (DE = 0)	5 (12.50)	0 (0.00)	24 (30.00)	9 (11.25)
Items with 4 NFDs (DE = 0)	-	-	4 (5.00)	7 (8.75)
Overall mean DE	56.67%	71.67%	48.44%	55.31%

Discussion

The end of phase examination in UPM MD program is a high-stake summative assessment at the end of preclinical and clinical phase. For preclinical phase examination, the results determine whether the preclinical students are eligible to progress to clinical phase, while the final year students need to pass the clinical phase examination to graduate. Therefore, valid assessment tools are needed to measure students' knowledge, skills and attitude in the examination. One of the tools used to test the 'knows' and 'knows how' in Miller's pyramid is with MCQ [7]. It is useful in measuring factual recall, but it can also test higher order of thinking skills such as application,

analysis, synthesis and evaluation of knowledge, which are important for medical graduates. Post-validation of test items using item analysis of PI, DI and DE is a simple, yet effective method to assess the validity of the test. In the present study, we analysed the MCQs from both preclinical and clinical phase examinations taken by two different cohorts of students. Each MCQ in preclinical phase examination has four options while clinical phase examination has five options.

Based on the findings, the mean PI in both examinations in both years were similar. This indicates that an increased number of options, five versus four options, does not have a significant impact on the difficulty level of the examination. This was supported by previous study by Schneid et al. [8] which found that there was no significant differences in the difficulty level among MCQs with three, four or five options. In contrary, Vegada et al. [9] found a slight decreased in the difficulty level when reducing the options from five to four, and the items became much easier when reducing the options to only three. They concluded that the items became easier with less number of options due to the increased probability of random guessing to select the correct answer.

Preclinical phase examination showed a consistent level of item difficulty for both years. Yet, half of the questions were 'very easy' and 'difficult'. These questions seem to be unsuitable to assess students in the high-stake examination as they were unable to discriminate between the good and the weak students. Hence, these questions should be revised, by changing either the vignette or the options. All difficult questions should be reviewed for its language and grammar, ambiguity and controversial statements [4]. Some previous studies demonstrated their PI as a percentage, which reflects the percentage of correct answers to the total responses [10–12]. An item with PI percentage between 30% to 70% is considered as acceptable, i.e. not too easy and not too difficult [11, 12]. Studies by Sim & Rasiah [10] and Rao et al. [11] showed a comparable mean PI similar to our present finding, ranging from 50 to 60%. It is evident that with a proper vetting and training in constructing MCQs led to such desirable findings. It was also

suggested that a continuous training and feedback should be given to the question authors so that the number of 'difficult' and 'very easy' questions can be reduced in the future

The mean DI for preclinical phase examination in both years, and 2017 clinical phase examination ranging from 0.25 to 0.31 which were considered as 'good'. An earlier study had shown that a comparable mean of DI proved that the quality of questions has been consistent along the years [4]. Nevertheless, the mean DI in clinical phase examination reduced significantly in 2018. This may be due to the high number (66%) of 'very easy' and 'difficult' questions in the examination. Consequently, about 53% of the questions were considered as 'poor' and were not able to discriminate between the good and weak students. A test item ideally should be able to pick out the 'good' students from the 'poor' ones, in which more 'good' students are able to answer the item as compared to the 'poor' students [5]. In the present study, some questions were found to have zero and negative DI. Zero DI means that the item was non-discriminating in which either all students were able to answer the item correctly, or an equal number 'good' and 'poor' students were able to answer correctly, or none of the 'good' and 'poor' students managed to correctly answer it. Negative DI indicates that more 'poor' students were able to answer the item correctly. In this study, one question was found to have zero DI and zero PI. These demonstrated an extremely low number of students who managed to answer it correctly and none of them were the 'good' and 'poor' students. We speculate the reasons for these were due to ambiguous framing of the questions, poor preparation of students or perhaps, wrong answer key was given [10, 13, 14]. One other question has zero DI and PI equals to 1 demonstrating that all students managed to answer it correctly, probably because it was too easy. Too difficult and too easy questions may contribute to the 'poor' questions based on the dome-shaped correlation between PI and DI [12, 15]. These questions were not useful and may reduce the validity of the test, therefore should be eliminated.

Preclinical phase examination in 2018 showed an increase in the number of 'excellent' and 'good' questions, and a decrease in 'poor' questions comparing to 2017. This proves that the preclinical lecturers have shown considerable improvement in constructing MCQs through continuous training and feedback. In contrast, less than 8% of the questions in 2018 clinical phase examination were considered as 'excellent' and there was an increased number of 'poor' questions compared to 2017. Several possible reasons to this finding were identified. Some of the clinical lecturers were new and probably were unfamiliar with the test format, while some clinical lecturers had never attended any training on test item construction. There should be a thorough and several levels of vetting by peers, who are content experts and non-content experts is also needed before the questions are used in an examination. With this analysis, feedback should be given to all authors for them to reflect upon and revise their questions accordingly.

The number of NFDs also affects the discrimination power of an item [11]. From this study, more than 20% of the MCQs for preclinical phase had no NFDs in both years. In fact, there was one third of the 2018 preclinical phase MCQs with three functioning distractors (DE = 100%). We identified that items with higher number of options in clinical phase examination tend to have higher NFDs in both years. Less than 10% of the MCQs for clinical phase examination had 100% DE. This shows that it was probably difficult for the authors to develop four equally plausible distractors. Preclinical phase examination in 2018 has no question with 0% DE; and it showed the highest overall mean DE as compared to other examinations. Preclinical lecturers have shown a significant improvement in developing MCQs with plausible distractors and avoiding NFDs between the years. The findings suggested that reducing the number of options may increase the credible of distractors in an item, although it may reduce its difficulty level.

A meta-analysis by Rodriguez [16] found that having three options in an item is adequate. Even though the difficulty level is lowered, but it is more discriminating and more reliable. This is supported by a more recent study which found that questions with even as low as three options

would still produce good reliability and less laborious to construct [17]. However, this means that students will have high chance (only 1 in 3) of correctly answering the item with random guessing. Royal & Dorman [18] highlighted that 3-option and 4-option MCQs had similar psychometric properties which means the former is equally effective as the latter. Therefore, the traditional MCQs with 4 options shall be maintained as more research needs to be done to better understand the effects of 3-option MCQs on guessing strategies and cut score determination decisions to avoid any unintended consequential validity [19].

The present study highlights some interesting findings. First, an increased number of options does not affect the difficulty level of the questions, however, it does have significant effect on their discrimination power. Questions with higher number of options tend to have lower DI and higher number of non-effective distractors. It is therefore suggested to standardize the number of options to only 4, in both preclinical and clinical phases of examination. Second, question authors from preclinical phase showed considerable improvement in constructing test items with plausible distractors and ideal difficulty level as compared to the authors from clinical phase. Lack of motivation and time constraints may be the possible challenges for the clinical authors to construct good quality items [20]. Despite the availability of the faculty's guidelines for constructing examination questions for reference, training and continuous follow-up and feedback to authors are important to decrease items flaws and improve item-writing skills. Institutional support for faculty development program is crucial to ensure reliable and valid assessment strategies especially for high-stake examinations.

The roles of vetting committee in medical schools have been described in literatures to evaluate the content, language and technical aspects of all questions [21, 22]. Vetting session should be more thorough to ensure the validity of test items by removing any flaws and making them as understandable and clear as possible [23]. Additionally, time as well as resources are needed to develop the assessment blueprint to ensure all items are aligned with the learning

objectives. A well-developed blueprint also corresponds with the depth of knowledge and level of difficulty of each content area.

Several limitations should be noted in this study. First, this study was confined to one educational setting, limiting its generalization. Any effort to infer the findings to other educational settings needs to be done with caution. Second, several other parameters such as internal consistency and correlation between PI and DI were not measured in this study. Lastly, some variables such as author's previous training on writing MCQs and students' characteristics were not controlled during the analysis which may affected the study findings.

Conclusion

The findings suggest standardizing the number of options to only four as it does not much affect the difficulty level of the questions but will improve the degree of the items to discriminate between high and low achievers. This will also ease the authors on writing MCQs with equally plausible distractors. More trainings are required for the authors, especially from clinical phase to improve the quality of the items as seen in preclinical authors. Feedback should be given to all authors after analysis for them to reflect and make improvement. Good quality items shall be stored in the question bank while the poor ones shall be discarded.

List of abbreviations

UPM: Universiti Putra Malaysia; MCQ: Multiple-choice questions; PI: Difficulty or passing index; DI: Discrimination index; DE: Distractor efficiency; NFD: Non-functional distractor; SD: Standard deviation

1 **Declarations**

2 **Ethics approval and consent to participate**

3 Not applicable.

4

5 **Consent for publication**

6 Not applicable.

7

8 **Availability of data and materials**

9 The data in the present study are available from the corresponding author on reasonable request.

10

11 **Competing interest**

12 The authors declare that they have no competing interests.

13

14 **Funding**

15 This study did not receive any specific grant from funding agencies.

16

17 **Authors' contributions**

18 SKA and RH contributed to the conceptualization and study design. Data analysis and
19 interpretation were performed by SKA, FI, PSJK and RH. SKA drafted the manuscript. FI and NFZ
20 contributed substantially to its revision. All authors read and approved the final manuscript.

Acknowledgement

The authors wish to thank the Dean and Deputy Dean (Academic) of Faculty of Medicine and Health Sciences, UPM for their support and guidance. Also, appreciations to the staff in Academic Unit of the faculty; Ms Siti Zuraida Shahardin, Ms Siti Nor Husaine Husin, Mr. Muhammad Hakimi Bin Suhaimi and Mr. Mohd Esham Husain for their valuable assistance in data collection and documentations.

References

1. Kar S, Lakshminarayanan S, Mahalakshmy T. Basic principles of constructing multiple choice questions. *Indian J Community Fam Med*. 2015;1:65–9.
2. Messick S. Validity. In: Linn R, editor. *Educational measurement*. Washington, DC: American Council on Education; 1989. p. 13–103.
3. Linn R, Gronlund N. *Measurement and Assessment in Teaching*. 8th ed. New Jersey: Prentice Hall; 2000.
4. Hassan S, Hod R. Use of item analysis to improve the quality of single best answer multiple choice question in summative assessment of undergraduate medical students in Malaysia. *Educ Med J*. 2017;9(3):33–43.
5. Abdul Rahim A. *What those numbers mean: A guide to item analysis*. Kota Bahru, Kelantan: KKMED Publications; 2010.
6. Ananthakrishnan N. Item analysis-validation and banking of MCQs. In: Sethuraman K, Kumar S, editors. *Medical Education Principles and Practice*. Pondicherry: JIPMER; 2000. p. 131–7.
7. Miller G. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9):S63–7.
8. Schneid S, Armour C, Park Y, Yudkowsky R, Bordage G. Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard setting. *Med Educ*. 2014;48(10):1020–7.
9. Vegada B, Shukla A, Khilnani A, Charan J, Desai C. Comparison between three option, four option and five option multiple choice question tests for quality parameters: A randomized study. *Indian J Pharmacol*. 2016;48(5):571–5.

- 1 10. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in
2 true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Ann*
3 *Acad Med Singapore*. 2006;35(2):67–71.
- 4 11. Rao C, Kishan Prasad H, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice
5 questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res*.
6 2016;2(4):201–4.
- 7 12. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at
8 the department of paediatrics, Arabian gulf university, Manama, Bahrain. *Sultan Qaboos*
9 *Univ Med J*. 2018;18(1):e68–74.
- 10 13. Hassan S, Amin RM, Bt. Mohd Amin Rebuan H, Thwe Aung MM. Item analysis, reliability
11 statistics and standard error of measurement to improve the quality and impact of multiple
12 choice questions in undergraduate medical education in Faculty of Medicine at UniSZA.
13 *Malaysian J Public Heal Med*. 2016;16(3):7–15.
- 14 14. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher-developed
15 test in educational measurement and evaluation. *Cogent Educ*. 2017;4(1):1301013.
- 16 15. Mitra N, Nagaraja H, Ponnudurai G, Judson J. The levels of difficulty and discrimination
17 indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary
18 summative tests. *Int eJournal Sci Med Educ*. 2009;3(1):2–7.
- 19 16. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80
20 years of research. *Educ Meas Issues Pract*. 2005;24(2):3–13.
- 21 17. Loudon C, Macias-Muñoz A. Item statistics derived from three option versions of multiple-
22 choice questions are usually as robust as four- or five-option versions: Implications for
23 exam design. *Adv Physiol Educ*. 2018;42(4):565–75.

- 1 18. Royal K, Dorman D. Comparing item performance on three- versus four-option multiple
2 choice questions in a veterinary toxicology course. *Vet Sci.* 2018;5(2):55.
- 3 19. Royal KD, Stockdale MR. The Impact of 3-Option Responses to Multiple-Choice Questions
4 on Guessing Strategies and Cut Score Determinations. *J Adv Med Educ Prof.*
5 2017;5(2):84–9.
- 6 20. Karthikeyan S, O'Connor E, Hu W. Barriers and facilitators to writing quality items for
7 medical school assessments - A scoping review. *BMC Med Educ.* 2019;19(1):123.
- 8 21. Gopalakrishnan S, Udayshankar PM. Question vetting: The process to ensure quality in
9 assessment of medical students. *J Clin Diagnostic Res.* 2014;8(9):XM01–XM03.
- 10 22. Hassan S, Simbak N, Mohd Yussof H. Structured vetting procedure of examination
11 questions in medical education in Faculty of Medicine at Universiti Sultan Zainal Abidin
12 Malaysia. *Malaysian J Public Heal Med.* 2016;16(3):29–37.
- 13 23. Wadi M. Question vetting: theory and practice. *Educ Med J.* 2012;4(1):e1–4.

14