Evaluating the Quality of Multiple Choice Question in Paediatric Dentistry Postgraduate Examinations

*Mawlood Kowash, Iyad Hussein, Manal Al Halabi

تقييم جودة أسئلة الإختيار من متعدد في إختبارات الدراسات العليا في طب أسنان الأطفال

مولود كواش، إياد حسين، منال الحلبي

ABSTRACT: *Objectives:* This study aimed to evaluate the quality of multiple choice question (MCQ) items in two postgraduate paediatric dentistry (PD) examinations by determining item writing flaws (IWFs), difficulty index (DI) and cognitive level. *Methods:* This study was conducted at Mohamed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE. Virtual platform-based summative versions of the general paediatric medicine (GPM) and prevention of oral diseases (POD) examinations administered during the second semester of the 2017–2018 academic year were used. Two PD faculty members independently reviewed each question to assess IWFs, DI and cognitive level. *Results:* A total of 185 single best answer MCQs with 4–5 options were analysed. Most of the questions (81%) required information recall, with the remainder (19%) requiring higher levels of thinking and data explanation. The most common errors among IWFs were the use of "except" or "not" in the lead-in, tricky or unfocussed stems and opportunities for students to use convergence strategies. There were more IWFs in the GPM than the POD examination, but this was not statistically significant (*P* = 0.105). The MCQs in the GPM and POD examination were considered easy since the mean DIs (89.1% ± 8.9% and 76.5% ± 7.9%, respectively) were more than 70%. *Conclusion:* Training is an essential element of adequate MCQ writing. A general comprehensive review of all programme's MCQs is needed to emphasise the importance of avoiding IWFs. A faculty development programme is recommended to improve question-writing skills in order to align examinations with programme learning outcomes and enhance the ability to measure student competency through questions requiring higher level thinking.

Keywords: Examination Question; Student; Educational Measurement; Discriminant Analysis; Pediatric Dentistry; United Arab Emirates.

الملخص: **الهدف:** تقييم جودة أسئلة الإختيار من متعدد في إختبارين في برنامج الدراسات العليا في طب أسنان الأطفال من خلال تحديد عيوب كتابة الأسئلة ومؤشر الصعوبة ومستوى الإدراك التمييزي للأسئلة. **الطريقة:** تم عمل هذه الدراسة في جامعة محمد بن راشد للطب والعلوم الصحية في دبي بدولة الإمارات العربية المتحدة وتضمنت تقييم جودة أسئلة الإختيار من متعدد لإمتحانين سابقين في الفصل الثاني من السنة الدراسية 2017–2018 تم أدائهما من خلال منصة الإنترنت، وذلك في مادة طب الأطفال العام ومادة الوقاية من أمراض الفم، قام بجمع البيانات إثنان من أعضاء هيئة التدريس كلٌ على حدة عن طريق تسجيل العناصر المطلوبة (عيوب كتابة الأسئلة ومؤشر الصعوبة ومستوى الإدراك التمييزي) لكل سؤال. **النتائج:** تمَ تحليل عدد 185 سؤال إختيار من متعدد ذات 4-5 خيارات وتحمل إجابة صحيحة واحدة، غالبية الأسئلة (81%) كانت من نوع إسترجاع المعلومات و الباقي (19%) من النوع التحليلي، بالنسبة لحجم العيوب في الأسئلة المؤلفة حسب الشيوع المستنتجة كالآتي: إستعمال (ماعدا أو لا يوجد) في السؤال، إستخدام مفردات مخادعة أو غير واضحة وإستخدام إستراتيجيَة المقاربة، وبشكل عام كان هناك عدد أكبر من العيوب في الأسئلة في إختبار الوقاية من أمراض الفم مقارنة بإختبار طب الأطفال العام لكن الفرق لم يكن كبيراً إحصائياً. (P = 0.105)حدد مستوى صعوبة الإختبارين بالسهل حسب متوسط مؤشر الصعوبة للأسئلة (أكثر من 70%) في إمتحان مادتي الوقاية من أمراض الفم وطب الأطفال العام (89.1% ± 8.9% و 76.5% ± 7.9% على التوالي). **الخلاصة:** التقليل من نسبة وجود العيوب في أسئلة الإختبارات لمستوى مقبول يتم من خلال تقييم الإختبارات وتدقيق المحتوى وتدريب الكادر التعليمي في مجال مهارات وضع أسئلة الإختبارات، كما نوصي بإضافة الأسئلة ذات المستوى الإدراكي العالي لكي تتوافق مع نتائج التعلم المطموحة وتدعم قابلية قياس كفاءة الطلبة.

الكلمات المفتاحية: أسئلة الأمتحان؛ الطالب؛ قياس التعليم؛ التحليل التمييزي؛ طبّ أسنان الأطفال؛ الإمارات العربية المتحدة.

Advances in Knowledge
- *The adequate utilisation of multiple choice questions (MCQ) can enhance educational outcomes in dentistry especially in the Middle East and Gulf Cooperative Council countries; however, more research and training in MCQ creation is needed.*
- *Various factors may be used to assess MCQ items based on their item writing flaws, difficulty index and cognitive level.*

Application to Patient Care
- *High quality and effective MCQ items serve as a well-known and often utilised method for evaluating and assessing students. MCQs can assist dental students in achieving an exceptional dental education.*

---

*Department of Paediatric Dentistry, Mohamed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
Corresponding Author's e-mails: mawlood.kowash@mbru.ac.ae and mkowash@gmail.com

An examination should evaluate clinical skills and not merely the ability to recall information.[1] In addition to evaluating a student, assessment tools govern the methods chosen by students during their learning process.[2] Scouller investigated the effect of evaluation methods on students' learning techniques and found that examinees were generally more likely to adopt a superficial learning style when the evaluation doctrine was based solely on recollection of facts. In comparison, students and trainees were more likely to implement a more in-depth approach to learning if the test questions required higher levels of analytical skills and cognitive abilities.[2] Several studies have reported that the assessment tool affects examinees' and trainees' chosen styles of learning.[3–5]

Multiple choice questions (MCQs) are a well-known and often utilised method for assessment and are used either individually or in combination with other forms of evaluation and assessment. The advantages of MCQs include their reliability and content validity and their ability to reduce reliance on skills related to writing and self-expression.[6] High quality and effective MCQs are suitable for quantifying knowledge and perceptions of a given subject; therefore, this method of examination should be construed as accurately assessing applied practice.[6] In addition, for MCQs to be of high quality and effective they must be free of item writing flaws (IWFs).[7]

Single best answer (SBA) MCQ items were the most common assessment used for evaluation in didactic courses at the Hamdan Bin Mohammed College of Dental Medicine and Mohammed Bin Rashid University of Medicine and Health Sciences (MBRU) in Dubai, UAE. In addition, recently in dentistry more emphasis has been placed on undergraduate assessments through MCQs.[8] Therefore, this study aimed to evaluate MCQ items' quality in two postgraduate paediatric dentistry (PD) examinations by determining MCQs' IWFs, difficulty index (DI) and cognitive levels.

Methods

This study assessed an existing pool of MCQs used in two end-of-semester examinations during the 2017–2018 academic year at MBRU. The target courses were PD postgraduate courses in general paediatric medicine (GPM) and the prevention of oral diseases (POD). Examinations were accepted as data sources if they contained MCQs of 4–5 items (one single correct option and 3–4 distractors) of SBA-type summative questions. Some true/false and extended matching questions were excluded. Of the four PD faculty who produced the MCQ items, two were formally trained in MCQ design and assessment by the Royal College of Surgeons of Edinburgh. They independently reviewed each question according to predefined criteria. When debatable questions were encountered, joint faculty agreements were made with the help of a subject expert. The cognitive levels of each question item were analysed using Buckwalter's criteria, which is a revision of Bloom's taxonomy.[10,11]

Each MCQ item was assigned to one of three cognitive levels. Level one included lower order thinking questions which required recall of information. Level two questions tested understanding and interpretation of data. Level three included higher order questions which tested the application of knowledge for solving a particular problem. A list of 14 commonly occurring IWF criteria were used to identify IWFs in each question.[7,12] The list of IWFs included the use of absolute terms and opportunities for students to use convergence strategy. In using this strategy, students are able to answer the question by recognising that the correct answer includes common elements of other options.

The basic structure of an ideal SBA was proposed by Case and Swanson.[7] An effective question consists of a stem, which ideally should be a context-rich clinical case scenario or vignette that encourages the application of knowledge to a clinical situation followed by a lead-in, which states a question or a requirement from a candidate [Figure 1]. Ideally the lead-in should not include "except" or "not". The answer options should include one correct answer as well as a number of distractors and be homogenous (e.g. all focusing on diagnosis, investigations, medications or treatment options), plausible, of an appropriate length and uncomplicated. Options should avoid the use of "all" or "none of
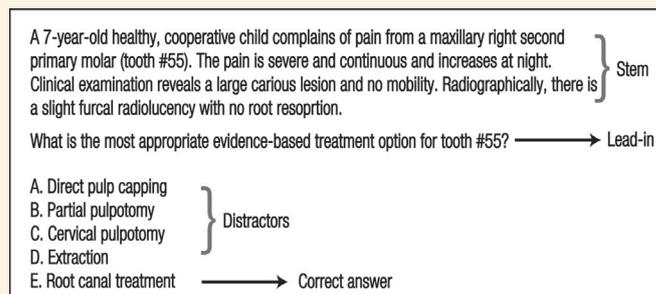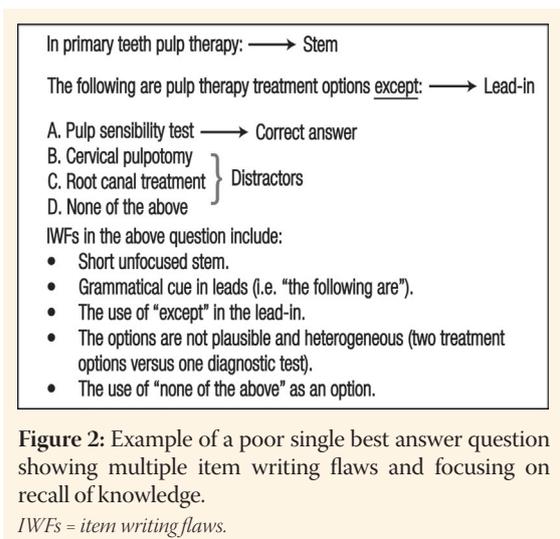

Figure 1: Anatomy of an effective single best answer question.

Figure 2: Example of a poor single best answer question showing multiple item writing flaws and focusing on recall of knowledge.

IWFs = item writing flaws.

Table 1: Distribution of cognitive levels and difficulty index in multiple choice questions from two examinations at Mohamed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates (N = 185)

Examination	Mean percentage ± SD	n (%)	
	Difficulty index*	Cognitive level	
		Level one	Level two
GPM[†]	89.1 ± 8.9	80 (84.2)	15 (15.8)
POD[‡]	76.5 ± 7.9	70 (77.8)	20 (22.2)
Total	-	150 (81.1)	35 (18.9)

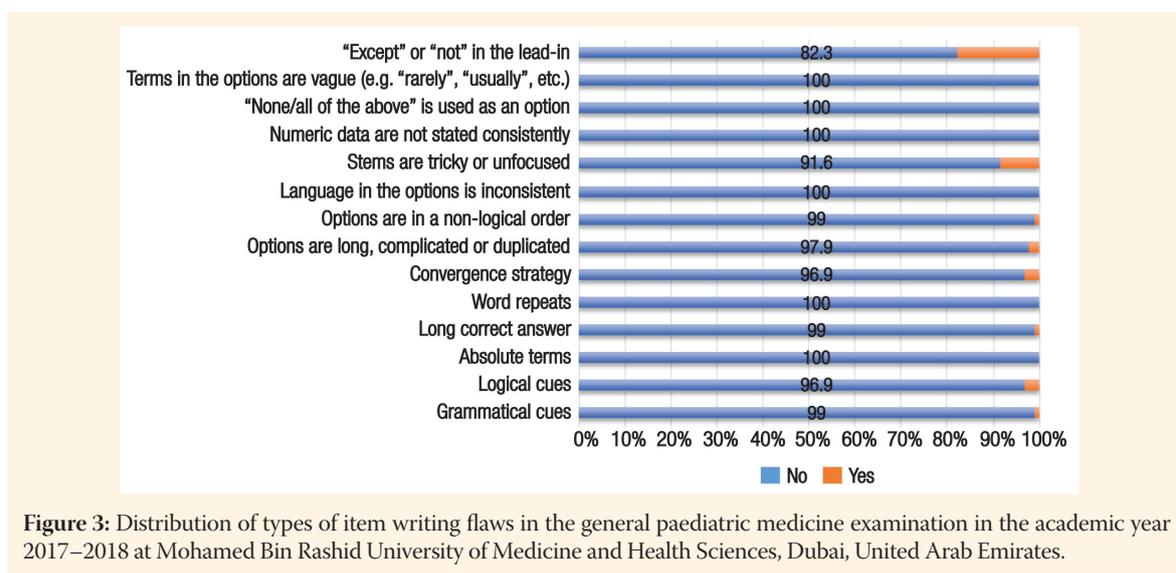*SD = standard deviation; GPM = general paediatric medicine; POD = prevention of oral diseases.*
Statistically significant at P <0.001. [†]n = 95. [‡]n = 90.

the above" or absolute terms such as "never". Options should also be absent of vague frequency terms such as "often" and "usually" and other IWFs. An example of an easy low-cognitive SBA question showing multiple IWFs is presented in Figure 2.

DI is defined as "the proportion of students who answered the item correctly, with the formula for the item-DI being p = c/n where, c is the number of students who selected the correct answer and n is the total number of respondents. The prop (proportion) value statistics ranges from 0 to 1"[13,14] The higher the prop value, the simpler the question was. Multiplying the prop value by 100 converts DI to a proportion. The prop value of the examinees who answered the question correctly could be classified as follows: <30% meant that the item was too difficult; between 30–70% meant that the item was good and acceptable; and a prop value >70% meant that the question was too easy and therefore unacceptable and in need of modification. The DI in an examination is defined as a measure of the effectiveness of an item in discriminating between high and low scorers.[13]
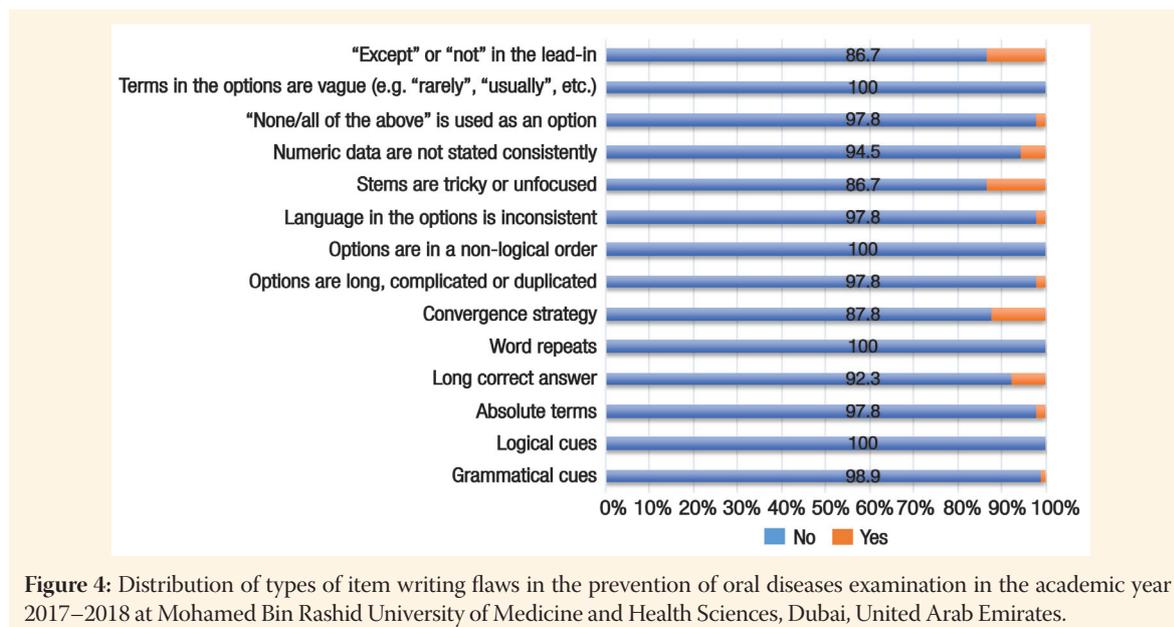
Descriptive statistics were used and statistical analysis was carried out using a pairwise t-test using Statistical Package for the Social Sciences (SPSS), version 20.0 (IBM Corp., Armonk, New York, USA). Statistical significance was set at $P <0.05$.

The MBRU Institutional Review Board approved an exemption as this research did not involve human subjects (MBRU-IRB-2018-010).

Results

A total of 185 SBA MCQs with 4–5 items (one correct option and 3–4 distractors) were analysed. The two PD faculty reviewers initially disagreed on 12 MCQ items (6.5%). The IWFs and/or cognitive levels of those questions were determined and agreed upon in a faculty meeting. Almost half of the questions (49.7%) had one or more IWFs in both examinations. The POD examination had more IWFs compared to the GPM examination (62.2% versus 37.9%). However, the difference was not statistically significant using a pairwise t-test ($P = 0.105$).

Figure 3: Distribution of types of item writing flaws in the general paediatric medicine examination in the academic year 2017–2018 at Mohamed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates.

Figure 4: Distribution of types of item writing flaws in the prevention of oral diseases examination in the academic year 2017–2018 at Mohamed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates.

Most MCQs (81.1%) required information recall (level one) while the remaining 18.9% required understanding and interpretation of data (level two). However, there was an absence of higher order thinking questions (level three) to test the application of knowledge. There was a significant difference in the mean DIs of GPM and POD MCQ items (89.1% ± 8.9% versus 76.5% ± 7.9%; *P* <0.001) [Table 1]. The most common IWFs in the general paediatric medicine [Figure 3] and the prevention of oral diseases [Figure 4] examinations were as follows respectively: the use of "except" or "not" in the lead-in (17.7% and 13.3%), tricky or unfocussed stems (8.4% and 13.3%) and opportunities for the use of the convergence strategy (3.1% and 12.2%).

Discussion

Effective MCQs are considered one of the best assessment tools available due to their validity, reliability, feasibility, educational impact and acceptability.[15] However, constructing standard and high-quality peer reviewed MCQ items requires training and practice.[16]

In the current study, the majority of questions (81.1%) tested recollection of isolated facts (level one) and the remainder (18.9%) tested comprehensive pooling of information (level two). None of the MCQs assessed the higher order cognition of applied practice and interpretation (level three). These findings were comparable with other studies which also found a focus on level one questions.[17–20]

Baig *et al.* evaluated 150 undergraduate pharmacology examination MCQs and found that most questions were at cognitive level one (76%) followed by level two (24%), with no questions written at level

three.[17] Tariq *et al.* found that the majority (60.47%) of the MCQs in an undergraduate pharmacology examination were at level one.[18] Tarrant and Ware evaluated an undergraduate nursing MCQ test and determined that >90% of the items were written at a lower cognition level.[19] Jozefowicz *et al.* studied the quality of MCQs in three American medical schools and reported an overall low quality of questions, most of which merely sought to assess students' recollection of basic dental information.[20] These studies and the high percentage of MCQs that tested low cognitive abilities in the present study could be attributed to the idea that MCQs were simpler to make, less time consuming and require less knowledge compared to higher order data synthesis items that demand expert input, time and training.[7,9] In the current study, the low cognitive levels of the MCQs can be attributed to the collection of examination questions from a recently established dental college with a limited question bank, which were created by various recently appointed faculty with inadequate training in question-writing. The effect of the latter was apparent when comparing the IWFs in the POD with the GPM examination (62.2% versus 37.9%). The newly appointed faculty contributed to constructing MCQs only in the POD test.

With proper training and adequate experience and resources, MCQs may be used to test students' higher cognitive skills.[21] For example, Dellinges and Curtis found that a one-hour MCQ training workshop for 24 dental faculty was effective in improving the quality of in-house MCQs when comparing pre-training and post-training MCQ-based scores in intervention and non-intervention groups.[22] Field *et al*'s study showed that constructing more challenging MCQs

involving problem-solving (level three) in clinical subjects was considered easier than basic science courses and was superior to other forms of questions.[8] In a study examining 50 MCQ items, Khan and Aljarallah reported that 60% of the items addressed the application of knowledge plane, 28% addressed recall of information (level one) but only 6% required interpretation of data (level two).[23]

In the present study, there were 92 IWFs (49.7%) in both postgraduate PD examinations. It is imperative to assess IWFs in MCQs because violations of accepted MCQ item-writing guidelines may affect examinee performance by making the item either easier or more difficult to answer.[24] Downing evaluated the quality of MCQ writing in four tests in the US and found that 46% of the items were classified as IWFs.[24] As a result of the IWFs, 10–15% of examinees who were categorised as "failures" would have been categorised as "pass" if flawed questions were excluded.[24] Tarrant and Ware studied the effect of IWFs on nursing examinees' achievements and reported that IWFs were frequent in high-stakes nursing assessments.[19] They did not penalise average examinees; however, high-performing examinees were probably more at risk than average students of being disadvantaged by IWFs.[19] The amount of IWFs in the current study may be attributable to an inadequately sized MCQ bank in this newly established college or inadequate formal question-writing training for the newly appointed faculty. Therefore, it is imperative that test creators reduce IWFs as they negatively affect difficulty and discrimination indices and might lead to a failure in achieving course learning objectives.[13,25]

The results of the present study showed more IWFs in the POD than the GPM examination (62.2% versus 37.9%); however, this difference was not statistically significant ($P = 0.105$). The most common IWFs in GPM and POD were the use of "except" or "not" in the lead-in (17.7% and 13.3%), tricky or unfocussed stems (8.4% and 13.3%) and convergence strategy (3.1% and 12.2%), respectively. Baig et al. reported a similar pattern of IWFs (46%) in their study; however, the four most frequent IWFs were the use of implausible distracters (30.43%), unfocused stems (27.54%), presenting unnecessary information in the stem (24.64%) or a negative stem (8.7%).[17] Downing also reported a comparable IWF proportion of 46%.[24] Khan and Aljarallah reported a lower IWF proportion (12%) on a problem-based learning examination.[23] In the present study, a higher proportion of IWFs can be interpreted in light of the Tarrant and Ware study. They stated that "MCQs written at lower cognitive levels are more likely to contain IWFs".[19] Tariq et al.

found fewer IWFs (28%) and also reported an increased proportion of level three questions in 150 pharmacology MCQs;[18] Baig et al.'s study of the same university determined 46% of the items had IWFs.[17] The authors of the aforementioned studies attributed the improvement to the in-house faculty's continuous medical education.

A post-validation item analysis of MCQ items should be conducted in order to evaluate correlations between item DI, discrimination and distraction effectiveness to determine whether questions should be reused, modified or discarded.[13] The present study evaluated a fairly large sample of MCQ items (N = 185) but in a small sample of postgraduate students; therefore, only the DI was analysed. The mean DI of the POD and GPM (76.5% ± 7.9% and 89.1% ± 8.9%) indicated that the MCQ items were easy (prop value >70%), especially in the GPM examination.[13,14] In comparison, Mukherjee and Lahiri reported a better DI mean prop value of 61.92% ± 25.1% in medical undergraduates.[26] Moreover, Mehta and Mokhasi reported various DI scores of which 62% of items were in an acceptable range (prop value 30–70%); 32% were too easy (prop value >70%) and 6% were too difficult (prop value >0.35).[27] Difficulty and discrimination indices are usually reciprocally related, but their relationship is often considered dome shaped and non-linear.[28] This finding suggests that questions with a high DI value discriminate poorly and vice-versa, except where the DI is either extremely high or low. One possible explanation for the high DI in the current sample is that the group consisted of only seven postgraduate residents with a high level of interest in the specialty and the examined topics.

In the current study, most MCQ items (81%) required knowledge recall (level one). Eliminating IWFs and using an examination template can improve cognition levels of MCQ test items.[25] Tarrant et al. challenged this idea and highlighted their belief that MCQs with IWFs were unlikely to alter question cognition.[29] Constructing MCQ items at higher cognition planes subsequently lead to the elimination of IWFs.[29]

In general, the quality of MCQ item writing in the two studied postgraduate PD examinations were comparable to the literature. As a result of this study, standardised question setting workshops were conducted. All future MCQ examinations will be subject to rigorous peer review, potentially improving the quality of MCQs by reducing/eliminating IWFs and constructing high cognitive level items with average difficulty and high discrimination. Open formal reflection, feedback and training regarding IWFs and MCQ analysis with faculty as well as students would help improve

learning outcomes. Periodic post-examination review of MCQ items available in the question bank would identify areas of potential weakness, thus helping to create an ideal item bank.

Conclusions

The most common IWFs in this study were the use of "except" or "not" in the lead-in, tricky or unfocussed stems and opportunities for students to use convergence strategy. Most MCQs were level one information recall items. A comprehensive review of the MCQ questions for all examinations in the program is needed with emphasis on avoiding IWFs. As a result of this study, a faculty development programme was recommended to improve the faculty's question writing skills and align examination questions with programme learning outcomes and enhance the ability of the questions to measure the competency of the students through questions that elicit higher order thinking.

References

1. Drew S. Perceptions of what helps learn and develop in education. Teaching High Educ 2001; 6:309–31. https://doi.org/10.1080/1356 2510120061197.

2. Scouller K. The influence of assessment method on students' learning approaches: Multiple-choice question examination versus assignment essay. High Educ 1998; 35:453–72. https://doi.org/10.10 23/A:1003196224280.

3. Trigwell K, Prosser M. Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. High Educ 1991; 22:251–66. https://doi.org/10.1 007/BF00132290.

4. Biggs J, Tang C. Teaching for Quality Learning at University. 4th ed. Philadelphia, USA: Open University Press, 2011. P. 103.

5. Reid WA, Duvall E, Evans P. Relationship between assessment results and approaches to learning and studying in year two medical students. Med Educ 2007; 41:754–62. https://doi.org/10.1111/j.136 5-2923.2007.02801.x.

6. Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO. The criteria and analysis of good multiple choice questions in a health professional setting. Saudi Med J 2005; 26:1505–10.

7. Case S, Swanson D. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia, USA: National Board of Medical Examiners, 2003. P. 31.

8. Field JC, Walmsley AD, Paganelli C, McLoughlin J, Szep S, Kavadella A, et al. The graduating European dentist: Contemporaneous methods of teaching, learning and assessment in dental undergraduate education. Eur J Dent Educ 2017; 21:28–35. https://doi.org/10.1111/eje.1 2312.

9. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educ 2012; 37:98–104. https://doi.org/10.1097/NNE.0b013e31825041d0.

10. Buckwalter JA, Schumacher R, Albright JP, Cooper RR. Use of an educational taxonomy for evaluation of cognitive performance. J Med Educ 1981; 56:115–21. https://doi.org/10.1097/00001888-198 102000-00006.

11. Huitt W. Bloom et al.'s taxonomy of the cognitive domain. From: www.edpsycinteractive.org/topics/cognition/bloom.pdf Accessed: Dec 2018.

12. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Measurement Educ 2002; 15:309–33. https://doi.org/10.1207/S1 5324818AME1503_5.

13. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore 2006; 35:67–71.

14. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos Univ Med J 2018; 18:e68–74. https://doi.org/10.18295/squmj.2018.18.01.011.

15. Dascalu CG, Enache AM, Mavru RB, Zegan G. Computer-based MCQ assessment for students in dental medicine–Advantages and drawbacks. Procedia Soc Behav Sci 2015; 187:22–7. https://doi.org/10.1016/j.sbspro.2015.03.005.

16. Botelho MG, Lam O, Watt RM, Leung D, Kember D. Evaluation of peer-generated MCQs to assess and support learning in a problem-based learning programme. Eur J Dent Educ 2018; 22:e358–63. https://doi.org/10.1111/eje.12304.

17. Baig M, Ali SK, Ali S, Huda H. Evaluation of multiple choice and short essay question items in basic medical sciences. Pak J Med Sci 2014; 30:3–6. https://doi.org/10.12669/pjms.301.4458.

18. Tariq S, Tariq S, Maqsood S, Jawed S, Baig M. Evaluation of cognitive levels and item writing flaws in medical pharmacology internal assessment examinations. Pak J Med Sci. 2017; 33:866–70. https://doi.org/10.12669/pjms.334.12887.

19. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008; 42:198–206. https://doi.org/10.111 1/j.1365-2923.2007.02957.x.

20. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med 2002; 77:156–61. https://doi.org/10.1097/00001888-20 0202000-00016.

21. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. BMC Med Educ 2007; 7:49. https://doi.org/10.1 186/1472-6920-7-49.

22. Dellinges MA, Curtis DA. Will a short training session improve multiple-choice item-writing quality by dental school faculty? A pilot study. J Dent Educ 2017; 81:948–55. https://doi.org/10.21815/ JDE.017.047.

23. Khan MU, Aljarallah BM. Evaluation of modified essay questions (MEQ) and multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate medical students. Int J Health Sci (Qassim) 2011; 5:39–43.

24. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract 2005; 10:133–43. https://doi.org/10.1007/s10459-004-4019-5.

25. Downing SM. Twelve steps for effective test development. In: Downing SM, Haladyna TM, Eds. Handbook of Test Development. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, Inc., 2006. Pp. 3–25.

26. Mukherjee P, Lahiri SK. Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR J Dental Med Sci 2015; 14:47–52. https://doi.org/10.9790/0853-141264752.

27. Mehta G, Mokhasi V. Item analysis of multiple choice questions-An assessment of the assessment tool. Int J Health Sci Res 2014; 4:197–202.

28. Menon AR, Kannambra PN. Item analysis to identify quality multiple choice questions. Nat J Lab Med 2017; 6:MO07–10. https://doi.org/10.7860/NJLM/2017/25690:2215.

29. Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. Nurse Educ Today 2006; 26:662–71. https://doi.org/10.1016/j.nedt.2006.07.006.