



ORIGINAL ARTICLE

Effectiveness of “fill in the blanks” over multiple choice questions in assessing final year dental undergraduates



R.M. Sumudu Himesha B Medawela^{a,*}, Dugganna Ralalage Dilini Lalanthi Ratnayake^b,
Wijeyapala Abesinghe Mudiyanse Udari Lakshika Abeyasinghe^b,
Ruwan Duminda Jayasinghe^a, Kosala Nirmalani Marambe^c

^a Department of Oral Medicine and Periodontology, Faculty of Dental Sciences, University of Peradeniya, Peradeniya, Sri Lanka

^b Department of Oral Pathology, Faculty of Dental Sciences, University of Peradeniya, Peradeniya, Sri Lanka

^c Medical Education Unit, Faculty of Medicine, University of Peradeniya, Peradeniya, Sri Lanka

Received 22 November 2016; accepted 14 March 2017

Available online 17 April 2017

KEYWORDS

Fill in the blanks;
Multiple choice
questions;
Guessing;
Effectiveness

Abstract

Background: Possibility of guessing in Multiple Choice questions (MCQ) when assessing undergraduates is considered a weakness. There are limited studies on the use of “Fill in the Blanks” (FIB) to overcome this issue.

Objective: To assess the effectiveness of FIB in MCQ for assessing final year dental undergraduates.

Methods and materials: A total of 134 final year dental undergraduates were randomly assigned to Group A and B. Group A was given a questionnaire with fifteen single best answer MCQ questions, and then the FIB questionnaire (which included the same questions in FIB form). At the same time Group B was given the FIB questionnaire initially, and then the MCQ questionnaire in the given period of time. The mean scores of the two groups were then compared.

Results: Group A obtained a mean score of 10.94 (SD $\pm$ 3.203) for MCQ, and 10.48 (SD $\pm$ 2.993) for FIB, whereas Group B obtained a mean score of 6.8 (SD $\pm$ 2.949) for FIB and 10.05 (SD $\pm$ 2.619) for MCQ. There was a statistically significant difference in the mean scores obtained for the two types of tests between Group A (P = .04) and Group B (P = .0001). The difference in the mean scores obtained for the FIB were statistically significant (P = .0001) between the groups, whereas the results were not statistically significant for MCQ (P = .127).

* Corresponding author.

E-mail addresses: sumuduhimesha@gmail.com, sumuduhimesha@pdn.ac.lk (R.M.S.H.B. Medawela).

PALABRAS CLAVE

Rellenar los espacios en blanco;
Preguntas de opción múltiple;
Valorar;
Eficacia

Conclusion: MCQ results revealed that the knowledge of the two groups was similar. The differences in the scores obtained for the two types of assessment tools suggest further research is needed to investigate the factors that led to the above observation.

© 2017 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

La eficacia de «rellenar los espacios en blanco» sobre las preguntas de opción múltiple en la evaluación de estudiantes de último año de odontología

Resumen

Antecedentes: Posibilidad de valorar si las preguntas de opción múltiple (MOM) son el punto débil cuando se evalúa a los estudiantes de último año. Hay estudios limitados sobre el uso de «rellenar los espacios en blanco» (REB) para superar este problema.

Objetivo: Evaluar la eficacia de REB sobre MOM en la evaluación de los estudiantes de último año de odontología.

Métodos y materiales: Se asignó aleatoriamente a un total de 134 estudiantes universitarios de último año de odontología a los grupos A y B. Al grupo A se le entregó un cuestionario con 15 MOM para seleccionar la mejor opción y luego un cuestionario REB (que incluía las mismas preguntas con planteamiento REB). Al mismo tiempo, al grupo B se le entregó el cuestionario REB inicialmente y luego el cuestionario MOM en el período de tiempo determinado. A continuación se compararon las notas medias de los 2 grupos.

Resultados: Un grupo obtuvo una puntuación media de 10,94 (DE $\pm$ 3,203) en MOM y de 10,48 (DE $\pm$ 2,993) en el REB, mientras que el grupo B obtuvo una puntuación media de 6,8 (DE $\pm$ 2,949) en el REB y de 10,05 (DE $\pm$ 2,619) en el MOM. Hubo una diferencia estadísticamente significativa entre las notas medias obtenidas de los 2 tipos de pruebas entre el grupo A ($p = 0,04$) y el grupo B ($p = 0,0001$). La diferencia en las notas medias obtenidas del REB fue estadísticamente significativa ($p = 0,0001$) entre los grupos, mientras que los resultados no fueron estadísticamente significativos en el MOM ($p = 0,127$).

Conclusión: Los resultados del MOM revelaron que el conocimiento de los 2 grupos era similar. Las diferencias en las notas obtenidas de los 2 tipos de instrumentos de evaluación sugieren que se necesita más investigación para analizar los factores que llevaron a la reflexión anterior. © 2017 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Assessment of students' competencies is fundamental in the undergraduate education. The term 'competence' is defined in medicine as "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individuals and communities being served".¹ A set of criteria for the evaluation of assessment methods are defined in the literature. These include reliability, validity, impact on future learning and practice, acceptability to learners and faculty and Cost to the individual trainee, the institution, and society at large.² One such method of assessment fulfilling these criteria not only provides an index for the evaluation of student's competencies but also provides a feed back to the students, which results in improvement of dental education.³ The method of educational assessment seems to affect student learning approaches and influence the performance of the students.

Use of Multiple choice questions (MCQ) is widespread in both formative and summative assessments. MCQ is a form of assessment in which respondents are asked to select the

best possible answer or answers out of the choices from a list. Multiple choice tests were developed initially by the psychologist Edward Thorndike (1874–1949). However the first all multiple choice large scale assessment was the assessment of the intelligence and more specifically the aptitudes of World War I military recruits in Army Alpha. There is a long standing criticism on validity of MCQ as it tests only the cognitive knowledge which lies on the lowest level of framework for assessing clinical competence proposed by famous psychologist George Millers in 1990.^{2,4} However the best answer MCQs used on common merit exam are scenario based and require application of knowledge and problem solving. Researches in medical education are in search of assessment options to overcome above issues arisen in assessment with MCQ. Extended matching question, two tier MCQs are some of them.

However "fill in the blanks" type questions, which has a structural similarity to MCQ (As both have short answer of one or few words) is more objective and overcome some of the disadvantages of MCQs such as possibility of guessing an answer. Guessing could be eliminated in fill in the blanks as they do not provide options for the student and student who

MCQ	Fill in the Blanks
The volcano known as Eyjafjallajokull , is closest to what Icelandic city? A. Reykjavic B. Kopavogur C. Hafnarfjorour D. Fjaroabyggo	The volcano known as Eyjafjallajokull , is closest to..... City in Iceland.

Figure 1 Example for a MCQ type of a question converted to a fill in the blanks type of the question.

knows the exact answer could get it correct. In MCQ version, a student who is unaware of the answer has a 25% (one out of four) chance of getting the correct answer with guessing. In contrast possibility of guessing is eliminated in the fill in the blanks version (Fig. 1).

However literature in utilization of the “fill in the blank” type of questions as a formal assessment method in tertiary education is or was found to be scarce and is certainly limited to primary education in assessing children. Therefore this study was carried out to assess the effectiveness of “Fill in the blanks” over multiple choice questions in assessing final year dental undergraduates of Faculty of Dental Sciences, University of Peradeniya as an initial step leading to exploration of a novel field in medical education.

Methods and materials

This study was an experimental study, carried out among 134 final year dental undergraduates in the Faculty of Dental Sciences, University of Peradeniya, Sri Lanka. A question paper, which included fifteen MCQs with a single best answer was used parallel with a paper which included the same fifteen questions in “fill in the blanks” form in order to assess the difference obtained in the scores. This arrangement enabled comparison of two student assessment tools.

MCQ questions were selected from the MCQ bank of the division of oral medicine and periodontology, Faculty of Dental Sciences whose reliability and validity had been assessed previously. Selected MCQ questions were then converted in to fill in the blank type of questions by a MCQ core group consisting of principal investigator, Professor in Oral Medicine, Consultant and Senior Lecturer in Oral Medicine, Language expert, Statistician with special interest in Medical Education. It was ensured that the meaning was retained in both types of questions. The questions were in the same order in both the test tools and each question was given a similar score in each of the test tools. The total mark scored by a student for the two tests were calculated separately.

Each of the questions was taken in to discussion among the members of the core group on the basis of clarity and content. The two tools were pre-tested by the principle investigator among 20 dental graduates who passed out recently and the questions were modified to achieve better clarity after the feedback of the participants of the pilot study. As an example two questions used in the study (In both MCQ format and FIB format) are shown in the following Fig. 2.

To create a safe environment, students were informed that the participation was not compulsory, and assured that their performance in the test was confidential and would not

Question number	MCQ	Fill in the blanks
04	Which one of the following terms is most appropriate to describe “Pain due to a stimulus which does not normally provoke pain?” 1. Allodynia 2. Causalgia 3. Hyperalgesia 4. Hyperesthesia 5. Hyperpathia	The most appropriate term to describe “Pain due to a stimulus which does not normally provoke pain” is _____.
12	Which one of the following would be the characteristic appearance of the sialogram performed on the parotid gland of a patient with “Sjogren syndrome” 1. Ball in hand appearance 2. Bush in winter appearance 3. Globular sialiectasis 4. Tree in winter appearance 5. Snow storm appearance	_____ appearance would be the characteristic appearance of the sialogram performed on the parotid gland of a patient with “Sjogren syndrome”.

Figure 2 Two examples in questionnaires (In MCQ format and FIB format) used in the study.

count in any way toward their course assessment. Students were reassured that they were not expected to perform beyond their ability as fourth year dental students.

Another questionnaire consisting of four questions to obtain the students feedback regarding the two types of tests was used following main questionnaire. The study sample consisting of final year dental students were randomly assigned to two groups (Group A and Group B) with the help of a computer generated random number. Instructions were given prior to the assessment. A number was allotted to a student randomly and the test was performed under that number in both the tools in order to maintain confidentiality.

Group A was given the MCQ questionnaire for the first 15 min and simultaneously Group B was given the fill in the blank questionnaire. In the second half of the assessment Group A was given the fill in the blank questionnaire to answer the questions within 15 min and simultaneously Group B was given the MCQ type questionnaire to answer. Students were strictly advised and monitored not to discuss the answers during the break.

During marking of the answer scripts they were marked by the principal investigator. Spelling mistakes were accepted to avoid confounding effect. Marked obtained for the each test by different groups were analyzed using Statistical package for the social sciences version 17.

Results

A total of 134 students were included in the study sample and among them 47 (35%) were males whereas 87 (65%) were females.

Total mean score obtained by the Group A for the MCQ and FIB was 10.94 ± 3.203 (mean $\pm$ SD) and 10.48 ± 2.993 (mean $\pm$ SD) respectively. The average scores obtained for the two types of tests among the students from the Group A were statistically significant ($t(68) = 2.089$, $P = 0.04$) at $P = 0.05$. Also the results obtained by the Group B was 10.05 ± 2.619 and 6.8 ± 2.949 for the MCQ type of questions and FIB test respectively and showed a statistically significant difference ($t(64) = 12.251$, $P = 0.0001$) at $P = 0.05$ level (Table 1).

Furthermore the average score achieved from the FIB test was lower among the students from the Group B compared to Group A which received the FIB type of questions as the

Table 1 Average scores obtained by two groups for each assessment method.

	MCQ (mean $\pm$ standard deviation)	Fill in the blanks (mean $\pm$ standard deviation)
Group A	10.94 ± 3.203^a	10.48 ± 2.993^b
Group B	10.05 ± 2.619^a	6.8 ± 2.949^b

^a $t(68) = 2.089$, $P = 0.04$ ($P < 0.05$).

^b $t(64) = 12.251$, $P = 0.0001$ ($P < 0.05$).

first method of assessment. This difference observed in the average scores obtained for the "fill in the blanks" type of test were statistically significant in between the Group A and B ($t(132) = 7.161$, $P = 0.0001$) at $P = 0.05$ level.

Additionally responses obtained for the questions in the feedback questionnaire are shown in Table 2.

Discussion

Several advantages and disadvantages of MCQ have been identified in the literature.³⁻⁸

Multiple choice questions have several advantages over other assessment methods. Importantly it is an effective assessment system in the assessment of Knowledge domain of the Millers pyramid. Also it requires less time to administer, creates a lower likelihood of teacher bias in the results. It is economical as it test knowledge quickly within large groups and it can be used to provide quick feedback. Since one can include many questions in a one-hour paper, could ensure content validity as well. That is to say different content areas could be tested in one paper in contrast to a few content areas in essay papers. Also reliability is ensured as answer variation is limited. With the newest technology MCQs can be automatically scored and can be analyzed with regard to difficulty and discrimination. Also they can be stored in banks of questions and re-used as required.

Considering the disadvantages of MCQ assessment most important disadvantage had been identified as it only assesses lowest levels (Knowledge and problem solving) in millers pyramid.⁹ Possible ambiguity in the examinee's interpretation of the item and the fact construction of MCQs take a lot of time are some other disadvantages. MCQs are less

Table 2 Responses obtained for the questions in the feedback questionnaire.

Question no.	Question	Answers	
		Yes %	No %
01	Have you ever encountered a "Fill in the blanks" type question during undergraduate program?	19	81
02	Do you feel the "Fill in the blanks" type questions challenged you more than a MCQ type question?	90	10
03	Do you feel the "Fill in the blanks" type questions are a good solution to avoid intelligent guessing in MCQ?	66	34
04	As a student do you feel you should have "Fill in the blanks" type questions in your undergraduate exams?	13	87

likely to test creativity but to a certain extent analysis could be tested with MCQ. There are different types geared at testing such levels.

In the context of assessing different skills MCQs are less likely to test literacy, or ability to analyze and also creativity, or unique thinking. Also this assessment method could encourage students to take a surface approach to learning. This is valid for factual recall type MCQs not the one currently proposed in Medicine. One of the most important disadvantages of MCQ is possibility of guessing an answer.¹⁰⁻¹²

However "fill in the blanks" type questions overcome some of the disadvantages of MCQ.

In this study the scores obtained for the MCQ by both groups were similar and the difference was insignificant. Hence we conclude the baseline scores revealed that the two groups which were randomly selected possessed a similar level of knowledge on the relevant discipline. Thus the significant difference observed in the scores for the 'FIB' as compared to MCQ by both groups displayed the challenge exerted by the former method of assessment. Further among the two groups those who obtained the MCQ first and FIB later had a better score for FIB than the other group. This finding could be attributed to the carry over effect of the clues gained by the MCQ test which is a known limitation in this type of assessment method.

Performance of students is poor with FIB though both groups have scored around same with MCQs. so the guessing factor is considerable. FIB is proposed as a comparable tool to MCQ, which takes off the guessing element and hence a more reliable assessment of student competency. However, whether the differences observed in this experimental study was merely due to intelligent guessing or any other limitation which is inherent to MCQ could not be concluded. To reveal on these facts properly designed experiments which has control of confounding variables should be conducted.

Nevertheless at the feedback session, 87% of the students displayed their objection to the use FIB type questions as an assessment tool in the undergrad program to assess their undergraduate academic performance while 90% of them perceived this as challenging than MCQ. In this respect it is questionable whether this technique fulfills the criteria of an acceptable assessment method. This may be partly due to the fact that students tend to prefer short cuts and they like recall type assessments. This method will require more of examiner time for construction may be less but marking has to be done manually. In case of selecting an assessment tool, using student popularity as a valid criterion is questionable in this regard and it has to be on the grounds of validity, reliability, feasibility rather than the popularity.

Therefore future exploration into this field is highly warranted in the field of medical education.

Conflict of interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

1. Ronald M Epstein. Assessment in medical education. *N Engl J Med*. 2007;356:387-96. <http://dx.doi.org/10.1056/NEJMr054784>.
2. van der Vleuten C. Validity of final examinations in undergraduate medical training. *BMJ*. 2000;321:1217-9. PubMed PMID: 11073517; PubMed Central PMCID: PMC1118966.
3. Vanderbilt AA, Feldman M, Wood IK. Assessment in undergraduate medical education: a review of course exams. *Med Educ Online*. 2013;18:1-5. <http://dx.doi.org/10.3402/meo.v18i0.20438>. PubMed PMID: 23469935; PubMed Central PMCID: PMC3591508.
4. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Med Teach*. 2004;26:709-12. Review. PubMed PMID: 15763874.
5. Mujeeb AM, Pardeshi ML, Ghongane BB. Comparative assessment of multiple choice questions versus short essay questions in pharmacology examinations. *Indian J Med Sci*. 2010;64:118-24. <http://dx.doi.org/10.4103/0019-5359.95934>. PubMed PMID: 22569324.
6. Anderson J. For multiple choice questions. *Med Teach*. 1979;1:37-42. <http://dx.doi.org/10.3109/01421597909010580>. PubMed PMID: 24483175.
7. Havyer RD, Nelson DR, Wingo MT, Comfere NI, Halvorsen AJ, McDonald FS, Reed DA. Addressing the interprofessional collaboration competencies of the association of American medical colleges: a systematic review of assessment instruments in undergraduate medical education. *Acad Med*. 2015 [Epubahead of print] PubMed PMID: 26703415.
8. Mathieu R, Nendaz, Ara Tekian. Assessment in problem-based learning medical schools: a literature review. *Teach Learn Med*. 1999;11:232-43. DOI:10.1207/S15328015TLM110408.
9. 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. PubMed PMID: 18045500; PubMed Central PMCID: PMC2148038.
10. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357:945-9. Review. PubMed PMID: 11289364.
11. Weller JM. Simulation in undergraduate medical education: bridging the gap between theory and practice. *Med Educ*. 2004;38:32-8. PubMed PMID: 14962024.
12. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. *Natl Med J India*. 2008;21:130-3. Review. PubMed PMID: 19004145.