

BRIEF REPORT

What steps are necessary to create written or web-based selected-response assessments?

Matt Morgan¹, Valérie Dory², Stuart Lubarsky³, Kieran Walsh^{4*}

¹BMJ Onexamination, Cardiff, United Kingdom; ²Undergraduate Medical Education & Centre for Medical Education, McGill University, Montreal, Canada; ³Department of Neurology, Montreal General Hospital, McGill Centre for Medical Education, Montreal, Canada; ⁴BMJ Learning, BMA House, Tavistock Square, London, United Kingdom

Abstract

Before we work out what constitutes an assessment's value for a given cost in medical education, we must first outline the steps necessary to create an assessment, and then assign a cost to each step. In this study we undertook the first phase of this process: we sought to work out all the steps necessary to create written selected-response assessments. First, the lead author created an initial list of potential steps for developing written assessments. This was then distributed to the other three authors. These authors independently added further steps to the list. The lead author incorporated the contributions of these others and created a second draft. This process was repeated until consensus was achieved amongst the study's authors. Next, the list was shared by means of an online questionnaire with 100 healthcare professionals with experience in medical education. The results of the authors' and healthcare professionals' thoughts and feedback on the steps, needed to create written assessment, are outlined below in full. We outlined the steps that are necessary to create written or web-based selected-response assessments.

Key Words: *Consensus; Feedback; Internet; Questionnaires*

Medical education is expensive [1]. Despite the progress that has been made in analysing what works in medical education, we know little about what constitutes value for a given cost in medical education [2]. This is particularly true of assessment in medical education. We know broadly what makes for a good assessment—it is an assessment that is valid, reliable, and fair and that has a positive impact on learners [3,4]. However these criteria have to be balanced against the costs of assessment. An assessment that is favourably 'balanced' in terms of its utility indices would cost little and score highly on these other indices of good assessment (i.e., validity and reliability, etc). This is true of any form of assessment—be it written assessment, objective structured clinical examination or

work-based assessments. However before we can work out what constitutes an assessment's value for a given cost, we must first outline the steps necessary to create an assessment, and then assign a cost to each step. In this study we undertook the first phase in this process: we sought to work out all the steps necessary to create written or web-based selected-response assessments [5]. Different assessment methods are associated with different steps even though broad categories of steps will likely overlap. We focused on written or web-based selected-response assessments first as they are widely used. The purpose of our study was to compile a comprehensive inventory of the steps potentially needed to create written or web-based selected-response assessments.

We used the following methodology to determine the steps necessary to create written or web-based selected-response assessments. First, the lead author created an initial list of potential requisite steps for developing written or web-based assessments in medical education. He then distributed this list to

*Corresponding email: kmwalsh@bmjgroup.com

Received: June 1, 2014; Accepted: November 9, 2014;

Published: November 8, 2014

This article is available from: <http://jeehp.org/>

the other three authors. These authors independently added further steps to the list. The lead author incorporated the others' contributions and created the second draft. This process was repeated until consensus was achieved amongst the study's authors. Next the list was shared by means of an online questionnaire with 100 healthcare professionals with experience in medical education. The questionnaire was shared by means of a listserv for medical educators. The listserv was Dr-Ed (<http://omerad.msu.edu/DR-ED/>) and has 2,237 members. Google forms were used to administer the questionnaire. These healthcare professionals came from 22 different specialties spanning primary and secondary care (respondent characteristics are outlined in Fig. 1). All respondents worked in medical education at undergraduate, postgraduate or continuing professional development levels. Respondents came from the following countries: Australia, Belgium, Canada, Colombia, India, Libya, Malaysia, Malta, New Zealand, Peru, Portugal, Singapore, UK, and the USA. They were asked to consider whether the list of steps put together by the authors was complete. Sixty four of the healthcare professionals felt that the list of steps of written assessments was complete. However 36 felt that it was incomplete. These submitted free text comments outlining the

steps that they felt were missing. The authors then incorporated these comments into the final version of the list. In summary we used a form of modified structured communication technique for developing a checklist for creating assessment. Ethical approval was obtained from the Commission d'Ethique Biomédicale, Université Catholique de Louvain, Faculté de Médecine, Montreal, Canada.

The results of the authors' and healthcare professionals' thoughts and feedback on the steps needed to create written assessment are outlined before, during, and after the test as well as at various points in the testing process were described. We have not included accommodation, travel or subsistence for any of the stakeholders. We have not included marketing for the exam. We assume that the curriculum and syllabus (including learning outcomes or objectives) have been developed and are available. We have not included remediation for those who have failed the exam. Results were as follows:

Before the test: Identifying and recruiting appropriate faculty; Training in test item writing; Training on exam software; Blueprinting (identification and selection of the objectives to be assessed and mapping of these objectives to the test items); Choosing test item format; Test item construction (including

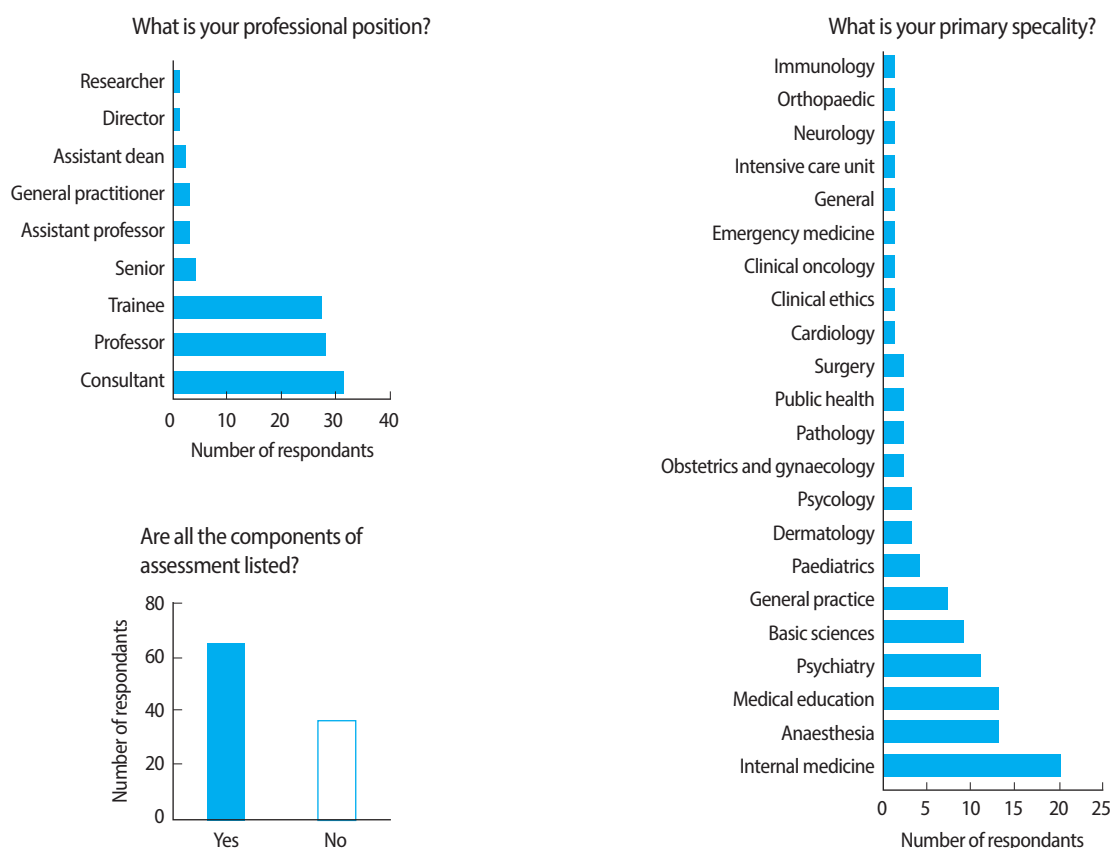


Fig. 1. Characteristics and answers of the survey respondents.

creation of explanatory feedback for correct and incorrect answers and reading and reviewing the literature and adding references); Multiple-stage improvement process (including peer review, editing and proof-reading); Creation or sourcing of visual or multimedia content (including licensing fees and cross checking media submitted and test items submitted); Selecting items from test item bank and coordinating test items from multiple instructors to make a coordinated exam; In computerized exams, loading test items into the testing system, tagging them for later retrieval or customization of tests, and adding supplemental files like audio or graphics files; Piloting; Item revision following piloting; Giving feedback to test item creators; and Knowing about which test items have been previously used in the exam.

During the test: Providing consumables (paper and printed forms); Invigilation (including training of invigilators); Providing hardware (e.g., computers); Providing software - if web based (including ensuring security of testing machines); Ensuring good user interface (design/layout of the screen and readability, quality of imaging, accessibility of testing resources such as lab values and calculators as needed) and testing environment (speed of computers, noise level); Sitting the exam (postgraduate learners); and Technology support.

After the test: Marking; Providing marking hardware and software; Giving/receiving feedback (including replying to individual students' concerns about test items); Dealing with appeals; Item analysis; Test analysis; Test item revision (or retirement); Feedback to faculty; Assurance processes to ensure probity and detect/eliminate cheating; Summarizing and reporting results to institutional stakeholders.

At various points in the testing process: Administration; Providing software (e.g., IT programmes, IT programmers' time, and electronic storage and transfer systems); Facilities (e.g., exam halls and storage); Standard setting; Senior inter-professional leadership and management team; and Evaluation and quality assurance of test and of the course itself in light of test results.

The added items after receiving the participants' further opinion are as follows: Identifying and recruiting appropriate faculty; Training on exam software; Creation of explanatory feedback for correct and incorrect answers; Cross checking media submitted and test items submitted; Selecting items from test item bank and coordinating test items from multiple instructors to make a coordinated exam; In computerized exams, loading test items into the testing system and tagging them for later retrieval; Item revision following piloting; Giving feedback to test item creators; Knowing about which test items have been previously used in the exam; Technology support; Dealing with appeals; Item analysis; Test analysis; Test item revision (or retirement); Assurance processes to ensure probity and detect/

eliminate cheating; and Summarizing and reporting results to institutional stakeholders.

In this study we have outlined the steps that are necessary to create written or web-based selected-response assessments. These included the steps before, during and after the test and at various points in the testing process. Although we feel that our list is comprehensive and includes major themes, we welcome comments from users who might feel that we have missed important steps. In creating the list we aimed to find a balance between creating something that was sufficiently detailed to be useful and usable, yet not so finely granular that it was too long, detailed or cumbersome. This study outlines the first phase in what is to be a multiphase research programme. Next we plan to assign a cost to each step of the process and then to conduct cost analyses of written assessments. The purpose of this first phase was to develop a standardized template for us and others to inventory different costs. It was also a starting point to design similar studies to define steps in other forms of assessment (e.g., open ended written formats or work-based assessments). According to Patricio et al's [6] systematic review of OSCE feasibility, cost reporting in the literature is highly variable. The strength of our study is that it includes the feedback of 100 respondents from different professional backgrounds and yet who all have experience in medical education. However there may be selection bias in terms of the responses. Interested persons are most likely to have responded which is appropriate for this sort of study; however, we may have missed some relevant experts.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

SUPPLEMENTARY MATERIAL

Audio recording of the abstract.

REFERENCES

1. Walsh K, editor. Cost effectiveness in medical education. Radcliffe: Abingdon; 2010.
2. Walsh K, Levin H, Jaye P, Gazzard J. Cost analyses approaches in medical education: there are no simple solutions. Med Educ. 2013; 47:962-968. <http://dx.doi.org/10.1111/medu.12214>
3. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach. 2011; 33:783-797. <http://dx.doi.org/10.3109/0142159X.2011.611022>
4. Norcini J, Anderson B, Bollala V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. Criteria for

- good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33:206-214. <http://dx.doi.org/10.3109/0142159X.2011.551559>
5. Hayes K, McCrorie P. The principles and best practice of question writing for postgraduate examinations. Best Pract Res Clin Obstet Gynaecol. 2010;24:783-794. <http://dx.doi.org/10.1016/j.bpobgyn.2010.04.008>
6. Patricio MF, Juliao M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach. 2013;35:503-514. <http://dx.doi.org/10.3109/0142159X.2013.774330>