

DATA NOTE

Open Access



Evaluation of MCQs from MOOCs for common item writing flaws

Eamon Costello^{1*} , Jane C. Holland² and Colette Kirwan¹

Abstract

Objective: There is a dearth of research into the quality of assessments based on Multiple Choice Question (MCQ) items in Massive Open Online Courses (MOOCs). This dataset was generated to determine whether MCQ item writing flaws existed in a selection of MOOC assessments, and to evaluate their prevalence if so. Hence, researchers reviewed MCQs from a sample of MOOCs, using an evaluation protocol derived from the medical health education literature, which has an extensive evidence-base with regard to writing quality MCQ items.

Data description: This dataset was collated from MCQ items in 18 MOOCs in the areas of medical health education, life sciences and computer science. Two researchers critically reviewed 204 questions using an evidence-based evaluation protocol. In the data presented, 50% of the MCQs (112) have one or more item writing flaw, while 28% of MCQs (57) contain two or more flaws. Thus, a majority of the MCQs in the dataset violate item-writing guidelines, which mirrors findings of previous research that examined rates of flaws in MCQs in traditional formal educational contexts.

Keywords: MOOCs, MCQs, Quality, Item writing, Tests, Education, Pedagogy

Objective

Despite increasing debate about the potential for Massive Open Online Courses (MOOCs) to contribute to formal, accredited qualifications, there is an absence of research into the quality of their assessments, including those based on Multiple Choice Question (MCQ) items. This provided the motivation to undertake an exploratory study of a selection of MOOCs to determine the existence and prevalence of item writing flaws in their MCQs. The full study and its findings are reported elsewhere [1, 2], but not the associated dataset provided here. We collected data from a sample of MOOCs (18) in the areas of medical health education, life sciences and computer science, and two researchers critically reviewed these 204 MCQ items using an evidence-based evaluation protocol derived from the medical health education literature [3]. Item writing flaws have been shown to compromise the reliability and validity of assessments and their outcomes, with inconsistent and unpredictable effects [4–6]. For instance, one study found that 33–46% of questions in a

series of science examinations were flawed, potentially incorrectly failing 10–15% of examinees who should have passed [4]. Another study found the converse, that some students who passed nursing examinations incorporating flawed items should arguably have failed [5]. Hence the data described here was generated with the objective of ascertaining whether such flaws pertain in MOOCs and to what degree.

Data description

The dataset incorporates assessments of human evaluators regarding MCQ quality in a selection of MOOCs. A set of 204 MCQs were collected by manually recording the questions, the options (or distractors) posed as potential answers and the actual correct answer from 18 MOOCs, and inputting these data into a spreadsheet. Two evaluators then independently reviewed the 204 MCQs, using an evaluation protocol adapted from Tarrent et al. [3] and which we include under workbook “Evaluation Instrument”. We did this through a Google Form. A unique id was assigned to each MCQ of the format $qi-j$, where i is the quiz to which the MCQ belongs and j is the number of the question in that quiz. We then pre-populated the Google form

*Correspondence: eamon.costello@dcu.ie

¹ Dublin City University, Glasnevin, Dublin 9, Ireland

Full list of author information is available at the end of the article



with a drop down menu of these identifiers. Each evaluator then selected the identifier of the question they were working on, and proceeded to fill in the form with their evaluations, which simplified and synchronised the workflow.

This led to the generation of the spreadsheet table which records evaluations for each of the 204 MCQs by each evaluator, considering 15 specific item flaws. For example, in the workbook “Evaluator 1 responses” cell A21 has the value “q11-7” which indicates it is the seventh question from quiz 11 in the data. In cell G21 the value “no” is recorded which tells us that evaluator 1 was of the opinion that this MCQ question did not contain plausible distractors. While the determination of what constitutes a plausible or implausible distractor is to some degree subjective (although may be supported by quantifiable data from statistical item analyses if available), the evaluators making this judgement were content experts. Implausible distractors can make it easy for student to guess the correct answer, as the correct option stands out as being the only obvious choice [7]. In the next workbook “Evaluator 2 responses” we can see that evaluator 2 recorded the same evaluation. The evaluators then compared their results, discussed any evaluations where their conclusions differed, and then agreed a final evaluation by consensus. These results are recorded in the “Combined Evaluations” worksheet.

Additional item writing flaws exist that can be identified without a human evaluator. These are: the number of possible correct options (1 is optimal); whether the correct option is the longest, as the longest is often the correct answer (these were calculated programmatically by counting the number of characters in each option); the number of options (3 or 4 are considered optimal [8]); whether “all of the above” or “none of the above” are options or not, as these violate best practice in item writing; and lastly, the position of the correct option (as research indicates option 3, or C, to most often be correct). These data points for each MCQ are recorded in the “Quantitative items” workbook. The workbook “Table 1 flaw prevalence” gives summary descriptive statistics of the raw results presented in the other workbooks. The final workbook, “MOOC platform and institutions”, lists the platform and the institution of each MOOC.

Limitations

Our sample of MOOCs and MCQs contained therein is a relatively small one. We had specific inclusion criteria and there were some practical limitations to our data collection. We examined only MOOCs delivered through the English language. We confined ourselves to topics in which the researchers’ had expertise i.e. focused in health sciences and computer science. We enrolled in MOOCs to take the quizzes which was a labour-intensive data gathering approach. Consequently, we sampled those courses that happened to be enrolling during the data collection window according to a convenience sample i.e. not a true random sample. However, we believe these data from this exploratory research to be potentially useful to others, given the dearth of research in this area, the growing importance of MOOC assessment (including MOOCs for formal credit) and the high prevalence of errors we found.

Most of our evaluations are qualitative in nature. To overcome this limitation two evaluators worked independently to perform their evaluations, and then met to compare results. Inter-rater reliability was calculated (Cohen’s Kappa score of 0.92) which indicated high agreement at this stage. Then the evaluators reviewed the items that were not in agreement and created revised combined scores.

We do not include the data on the questions in the MOOC quizzes themselves as they are copyright of the MOOC platform providers. However, we are happy to share further aspects of these data upon request to interested researchers.

Abbreviations

MOOC: Massive Open Online Course; MCQ: Multiple Choice Question.

Authors’ contributions

EC designed the study. EC and JCH conducted the research literature review. EC and CK collected the data. The EC and CK conducted the evaluations. EC conducted the analysis. All authors read and approved the final manuscript.

Author details

¹ Dublin City University, Glasnevin, Dublin 9, Ireland. ² RCSI Dept of Anatomy, 123 St Stephens Green, Dublin 2, Ireland.

Acknowledgements

The support of the Open Education Unit and the National Institute for Digital Learning in Dublin City University is acknowledged.

Competing interests

The authors declare that they have no competing interests.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Dataset MOOCs MCQs study v2.xlsx—3.0 MB	Excel file	Zenodo https://doi.org/10.5281/zenodo.1487858

Availability of data materials

The data described in this data note can be freely and openly accessed on Zenodo at <https://doi.org/10.5281/zenodo.1487858> [9].

Please see Table 1 for details and links to the data.

Consent for publication

Not applicable.

Ethics approval and consent to participate

No human subjects were part of this study and permission was thus not required according to the Institutional Review Board guidelines of author one.

Funding

The authors declare no source of funding.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 October 2018 Accepted: 26 November 2018

Published online: 03 December 2018

References

1. Costello, E., Brown, M., & Holland, J. (2016). What Questions are MOOCs asking? An Evidence-Based Investigation. Proc. European Stakeholder Summit (EMOOCs), Graz, Austria, 211–221.
2. Costello E, Holland J, Kirwan C. The future of online testing and assessment: question quality in MOOCs. *Int J Educ Technol Higher Educ*. 2018;15(1):42. <https://doi.org/10.1186/s41239-018-0124-z>.
3. 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 Pract*. 2006;26(8):662–71. <https://doi.org/10.1016/j.nedt.2006.07.006>.
4. 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*. 2005;10(2):133–43. <https://doi.org/10.1007/s10459-004-4019-5>.
5. 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(2):198–206. <https://doi.org/10.1111/j.1365-2923.2007.02957.x>.
6. Pais J, Silva A, Guimarães B, Povo A, Coelho E, Silva-Pereira F, Lourinho I, Ferreira MA, Severo M. Do item-writing flaws reduce examinations psychometric quality? *BMC Res Notes*. 2016;9(1):399. <https://doi.org/10.1186/s13104-016-2202-4>.
7. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners; 2002.
8. Rodriguez MC. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educ Meas*. 2005;24(2):3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>.
9. Costello E, Kirwan C. Dataset on an evaluation of MCQs from MOOCs for common item writing flaws [Data set]. Zenodo. 2018. <https://doi.org/10.5281/zenodo.1453420>.

Data citation

9. Costello E, Kirwan C. Dataset on an evaluation of MCQs from MOOCs for common item writing flaws [Data set]. Zenodo. 2018. <https://doi.org/10.5281/zenodo.1453420>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

