Physical and Engineering Sciences in Medicine (2022) 45:273–278
https://doi.org/10.1007/s13246-021-01088-x

SCIENTIFIC NOTE

Assessment of biomedical engineering knowledge using true–false questions

Manish Sreenivasa[1] · **Lucy Armitage**[1] · **Winson C. C. Lee**[1]

Received: 1 August 2021 / Accepted: 3 December 2021 / Published online: 20 January 2022

Abstract

The COVID-19 pandemic has caused a shift from on-campus to remote online examinations, which are usually difficult to invigilate. Meanwhile, closed-ended question formats, such as true–false (TF), are particularly suited to these examination conditions, as they allow automatic marking by computer software. While previous studies have reported the score characteristics in TF questions in conventional supervised examinations, this study investigates the efficacy of using TF questions in online, unsupervised examinations at the undergraduate level of Biomedical Engineering. We examine the TF and other question-type scores of 57 students across three examinations held in 2020 under online, unsupervised conditions. Our analysis shows significantly larger coefficient of variance (CV) in scores in TF questions (42.7%) than other question types (22.3%). The high CV in TF questions may be explained by different answering strategies among students, with $13.3 \pm 17.2\%$ of TF questions left unanswered (zero marks) and $16.4 \pm 11.5\%$ of TF questions guessed incorrectly (negative marks awarded). In unsupervised, open-book examination where sharing of answers among students is a potential risk; questions that induce a larger variation in responses may be desirable to differentiate among students. We also observed a significant relationship ($r = 0.64$, $p < 0.05$) between TF scores and the overall subject scores, indicating that TF questions are an effective predictor of overall student performance. Our results from this initial analysis suggests that TF questions are useful for assessing biomedical-theme content in online, unsupervised examinations, and are encouraging for their ongoing use in future assessments.

Keywords Proctoring · Remote learning · Student outcomes · Closed-ended questions

Introduction

The rapid onset of the COVID-19 pandemic in the year of 2020 and associated restrictions necessitated a paradigm shift in how we teach and assess student knowledge. The move to online and remote teaching methods has meant that both students and teachers have had to quickly adapt to new ways of delivering content and holding examinations. For example, in Australia, within a span of a few months a vast majority of universities moved graduate and undergraduate teaching to the online space using technologies such as Zoom, Webex etc. [1]. In this new environment, assessing student knowledge effectively via online examinations has been a particularly challenging task.

Online examinations tend to be open-book, allowing students to freely access textbooks and online resources, and are taken remotely (i.e. in a non-university setting). This raises some important issues in designing questions that effectively gauge individual student knowledge, such as the relative difficulty of the questions and the limited possibility of invigilation. While some online supervision (or *proctoring*) solutions have sprung up, there remain legitimate concerns regarding student privacy and the efficacy of the method [2, 3]. The unfettered access to information raises concerns regarding academic misconduct such as cheating [4–6], and a possibly reduced reliability of the scores [7–9].

Online examinations usually require the use of learning management systems (LMS) (e.g. Moodle, Blackboard, Canvas) and convert assessments traditionally held in-person to be instead held via the LMS. Most LMS systems allow assessment developers to incorporate a combination

✉ Winson C. C. Lee
ccwlee@uow.edu.au

[1] School of Mechanical, Materials, Mechatronic and Biomedical Engineering, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

of closed-ended formats, such as multiple-choice (MC) and true–false (TF) questions, and open-ended formats, such as free response (FR) questions. While using a spectrum of question types allows a better understanding of whether students have grasped concepts as well as facts [10], closed-ended formats are especially suited for the online environment and automatic grading. These formats have been used in a wide variety of fields and levels of study [11, 12]. Research has shown that closed-ended formats allow assessors to quickly gauge student knowledge over a large content area [13, 14]. In particular, multiple true–false (TF) format was found to accurately identify student competencies as well as gaps in knowledge [15, 16]. As students have a 50% chance of getting the answer correct in TF questions, negative marking for incorrect answers may be implemented to deter guessing.

While previous studies have reported the score characteristics in TF questions in proctored examinations [10, 13–18], there is little information on how this question format fares in online, open-book and unsupervised examinations. With the general move to online, remote teaching and the possibility of pandemic-related restrictions continuing in the future, understanding how such closed-ended formats perform in these new environments is important.

This pilot study examines the TF format question and statistically analyses the scores in online examinations. It adds to the literature by revealing in unsupervised open-book examinations (1) the differences in scores and score variations between TF and FR questions, (2) the correlation between the scores obtained in TF questions and the overall examination, and (3) whether the majority of marks losses in TF questions are due to negative marking for wrong answers or leaving questions unanswered. This study focuses on the field of Biomedical Engineering at the intermediate undergraduate level (2nd and 3rd year of a 4-year degree).

Methods

Students and academic subjects

We focus our study on Biomedical Engineering subjects run at the University of Wollongong (UOW), Australia in the academic year of 2020 (March–December). Three of the nine biomedical subjects in this time featured TF questions in the final examinations. The analysis involved 57 students, who were enrolled in the Bachelor of Engineering (Honours) in Biomedical Engineering. Students at this stage (2nd and 3rd year) have already received instructions on basic engineering concepts in the 1st year. In the 2nd and 3rd years, the subjects progressively specialise in topics arising from medicine, biology and engineering. This study received approval for the analysis of the examination scores of these students from the Human Research Ethics Committee at UOW (Ethics number: 2021/116).

The three subjects that were analysed were BMEG201: Biomedical Instrumentation and Design (24 students), BMEG301: Mechanics of Biomedical Systems (13 students), and BMEG303: Biomechanical Basis of Human Movement (20 students). These subjects covered the use of principles of mechanical and electronic engineering in addressing human-health issues. Due to the COVID-19 pandemic, all lectures of these subjects were conducted online in 2020.

Examination question types

Examinations featured a combination of MC, TF and FR type questions. The TF questions required students to decide if a statement was true or false. No calculations were required in these questions. When scoring the TF questions, one mark was awarded for a correct answer, zero marks if the student did not record a response, and a negative one mark if a wrong answer was selected. The minimum score for the TF section was zero (i.e. even if students got more incorrect answers than correct, their mark could never drop below zero for that section of the examination). All students were provided with some prior experience attempting the same question type before the examinations to have some familiarization with the format.

Some TF questions involved purely one statement. For example, “An X-ray image intensifier magnifies the dosage of x-rays to enable brighter images. True or False?”. Alternatively, some questions included a graph or a table providing further information for students to consider and then decide on the appropriate response (e.g. Fig. 1).

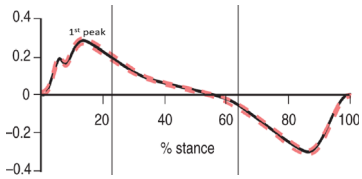
In each examination, students also attempted FR questions which required one or a combination of (1) calculations, (2) using engineering principles to explain Biomedical phenomenon, and/or (3) device and experimental designs. Standardized marking schemes were used to grade the FR answers. In addition, BMEG201 involved Multiple-Choice (MC) questions which required students to choose one of four choices which best address the question. No negative marking was applied for MC and FR question types. Table 1 shows the question types and their percentage weighting towards the final total examination scores in each of the three subjects.

Examination arrangement

All examinations were conducted online because of the COVID-19 pandemic and the associated shift to remote, online teaching at UOW. Students completed the examination remotely via Moodle, an open-source online LMS used at UOW. Students were required to submit their answers within a specified duration as listed in Table 1. Students

Fig. 1 Sample TF question with a figure from the BMEG303 examination

The following graph shows ground reaction force against percentage of gait cycle (y-axis indicates body-weight normalized forces).



The first peak as indicated in the figure is related to an anteriorly directed force. True or False?"

Table 1 Details of the three subjects and the question types within each examination

Subject codes	Question types and percentage of total examination scores (in brackets)		Examination durations (h)
BMEG201	10 True–false questions (25%)	10 MC* questions (25%), 5 Free-response questions (50%)	2
BMEG301	10 True–false questions (20%)	6 Free-response questions (80%)	2
BMEG303	20 True–false questions (25%)	5 Free-response questions (75%)	3

BMEG201 was a 2nd year subject with the other two being 3rd year biomedical engineering subjects

*Multiple-choice (MC) questions which involved 4 choices for each question

wrote their answers on their own piece of paper, scanned using a smartphone or a scanner and submitted it via Moodle. All three examinations were unsupervised with students completing their examinations at a location of their choosing. The examinations were open-book, i.e. students were allowed to use books and internet resources when attempting the examination questions. Students were required to read and accept an honour code at the start of the examination. The honour code stated that by commencing the examination students agreed that they would complete the examination without any aids from any other individual or group and that the violation of this agreement would lead to university academic misconduct procedures. The examination requirements and regulations were communicated to students at least 7 days prior to the start of examination.

Scores analysis

Student performance was manually marked by the subject lecturers. Statistical analyses were performed on the examination scores using IBM SPSS Statistics (v. 26) and Microsoft Excel (v. 2016). Means and standard deviations were computed for TF questions and all other question types. Coefficient of variance (CV), defined as the standard deviation divided by the mean, was then calculated to allow comparison of score variations among question types which had different mean scores. A paired t-test was performed to investigate if significance differences existed (1) in percentage scores between TF questions and all other question types and (2) between percentage of number of TF questions with wrong answers and percentage of such questions which were unanswered. Significance level was set at $p = 0.05$. Normality of data was tested and confirmed using the Shapiro–Wilk test. Correlation coefficient (r) was computed to study the strength of the relationship between TF questions scores and overall subject scores. Note that the overall subject scores included other assessments such as reports, presentations, quizzes etc., and were used here as a general indicator of student performance in the subject. The same statistical analyses were performed after pooling the scores from all three subjects together. Data used in these analyses may be accessed as supplementary data at this article.

Results

We observed that all three subjects had higher CV in TF questions (28.9–46.1%) than other question types (17.1–24.4%). Moderate to good associations with statistical significance ($r = 0.68$ and 0.7; $p < 0.05$) were found between TF question scores and overall subject scores in two subjects (Table 2). TF question scores were significantly lower ($p < 0.05$) than other question types in two subjects. Table 2 summarizes the results from the statistical analysis for each individual subject. Further analysis indicated that in BMEG201 the CV in TF questions was 46.1% compared to 11.0% in MC questions. Interestingly, no significant differences were found in each of the three subjects between the percentage of wrongly answered TF questions (10.8–18.5%) and the percentage of unanswered TF questions (11.5–17%) (Table 3).

Table 2 Score characteristics in each academic subject

Academic subjects	True–false questions		Other question types		Significant differences[a]?	Correlation coefficient[b]
	Percentage scores (%)	CV (%)	Percentage scores (%)	CV (%)		
BMEG201	53.3 ± 24.6	46.1	71.2 ± 12.3	17.2	Yes, $p < 0.05$	0.68*
BMEG301	66.9 ± 19.3	28.9	62.1 ± 15.2	24.4	No, $p > 0.05$	0.7*
BMEG303	46.0 ± 20.2	44.0	64.5 ± 11.0	17.1	Yes, $p < 0.05$	0.37
All three subjects	53.9 ± 23.0	42.7	63.6 ± 14.2	22.3	Yes, $p < 0.05$	0.64*

*Indicates a significant relationship against $p < 0.05$

[a]Differences in percentage scores between true–false questions and other question types

[b]Relationship between success rates in true–false questions and overall subject scores

Table 3 Percentages of TF questions with wrong answers and those left unanswered

Academic subjects	% questions with wrong answers	% unanswered questions	Significant differences?
BMEG201	17.7 ± 12.6	12.3 ± 18.7	No, $p > 0.05$
BMEG301	10.8 ± 9.5	11.5 ± 14.6	No, $p > 0.05$
BMEG303	18.5 ± 10.4	17.0 ± 17.7	No, $p > 0.05$
All three subjects	16.4 ± 11.5	13.3 ± 17.2	No, $p > 0.05$

Pooling the scores from all three subjects together, it was found that the CV for the full cohort in TF questions was 42.7% compared to 22.3% for other question types. Average percentage scores in TF questions (53.9 ± 23.0%) was significantly lower than other question types (63.6 ± 14.2%), with $p < 0.05$. There were 16.4 ± 11.5% TF questions with wrong answers (which attracted mark deductions), compared to 13.3 ± 17.2% of unanswered TF questions. There was no significant difference ($p > 0.05$) between number of wrong answered and unanswered TF questions. A significant correlation ($r = 0.64$; $p < 0.05$) was found between TF questions and overall subject scores.

Discussion

Our results clearly indicate that the TF question type elicit a larger variation in scores than other question types combined. The large variation may be due to the negative marking applied to the incorrect TF answers. It is likely that the potential for losing marks may have led to students employing different answering strategies based on their confidence with the material and individual appetite for risk. Negative marking has a mixed consensus in the literature. It has been shown to discourage guessing, but also to increase examinations-related anxiety for some students [17]. It has also been used as an incentive for students to identify areas where there is a lack of knowledge in disciplines where risk averse decision making is an important skill [19]. Our results show no statistically significant differences between percentages of questions left unanswered (13.3 ± 17.2%) and those answered incorrectly (16.4 ± 11.5%). As wrong answers attracted a negative mark, these results indicate that more scores were lost in TF questions due to incorrect answers. In the one subject where both MC and TF questions were presented in the same examination, CV was much lower in MC questions without negative marking (11.0%) than TF questions (46.1%). The large standard deviations suggest that students used different answering strategies (guessing an answer or leaving the question unanswered) when attempting questions they were not confident in answering. These results align with literature findings that the TF format of questions may be useful in identifying areas where students are less confident with the content [15, 16]. TF questions appeared to be good at identifying common misconception among students. For example, there was one question asking students to identify if the statement "The ground electrode for electromyography (EMG) recording adds electrical safety" was True or False. Many students (29.6%) mistakenly chose False, as they confused this with concepts related to the grounding in electrical plugs, leading to a larger CV in this question. Students gained good learning experience, as we reviewed the common misconceptions identified from the TF questions. In general feedback gathered after examinations, students agreed that the section containing TF questions was the most difficult, as they tended to lose more marks there. So far, we have not received any comments challenging the design of the TF questions.

In the context of an online, unsupervised, open-book examination where sharing of answers among students is a risk, questions that induce a larger variation in responses may be desirable to differentiate between students. Large

variations in scores with online exams have been reported previously in the literature [8]. Hollister et al. propose three reasons for this variation: the online location, learning style preferences and the opportunity for cheating with reduced proctoring. While difficult to test conclusively, it may be possible that the negative marking also resulted in students being less willing to share and use responses from others in the class when the risk of a loss of marks for an incorrect response is present.

The significant relationship between overall subject scores and scores in TF question responses suggests that these questions were a good predictor of overall student performance. Since the CV was also larger, it is possible that the other questions did not differentiate as much between students, meaning that the TF responses were the main point of difference for students within the examination. In particular, the loss of marks for incorrect responses resulted in an increase in the variation between students. In the context of online exams where invigilation and ensuring students do not communicate or share responses is difficult or impossible, this point of differentiation may be particularly desirable.

Our results are encouraging and relevant as university education and assessment continues to be conducted remotely. However, we also identify the following limitations to our present analysis and scope. First, we have limited retrospective data that was analysed as part of this pilot study. As progressively more students pass through the biomedical stream at UOW, we may expect to uncover additional insights with the inclusion of more data. Second, there was no comparison with data from courses using true false questions without negative marking to see whether it was the question style or the marking approach that was the bigger effect. Third, it is possible that a poor format or quality of the TF questions could lead to a higher CV. However, we note that no such feedback was received from any of the cohorts tested. Additionally, that good correlation between TF scores and overall subject scores, indicate that, at least, the TF questions did not introduce arbitrary variations in the scores. Finally, TF questions may not be universally appropriate to assess different skills and disciplines. The format lent itself well for some of the biomedical themes tested in this study. However, it should be noted that the more numerical, mechanics and electronics related questions were better suited to the FR format due to requirements such as providing detailed calculations, free-body diagrams etc. In this context, it is also notable that in this study the TF questions contributed up to 25% of the grade, and that we continued to use the other question types (including the FR format) as a valuable instrument for the assessment of engineering knowledge. While many fields in life sciences may be also be assessed using the TF format, it is likely that the TF format will serve as an adjunct rather than absolute solution to online assessment difficulties.

Conclusions

Our study has shown that in a cohort of biomedical engineering students, there was an increase in the variation of responses using true–false questions with negative marking when compared to multiple choice and free response type questions. Higher variation in responses in unsupervised, open-book examinations may be desirable, as sharing of answers among students is a risk in such examination conditions. The high variation in scores in TF questions may be due to the use of different answering strategies among students, with a similar number of students incorrectly guessing TF responses or leaving them unanswered. There was also a statistically significant relationship between total mark for the TF questions and overall subject mark. Our results support the use of the TF format to be used as a tool for students and educators to identify areas where there is a lack of knowledge and highlight its usefulness in online, unsupervised biomedical-themed examinations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13246-021-01088-x.

Acknowledgements The Authors would like to thank the teams at the Learning, Teaching & Curriculum (LTC) and the Student Services Division (SSD) at the University of Wollongong, Australia for their support towards data usage in this study.

Author contributions WL conceived the study and performed the data analysis. MS, LA and WL evaluated the results and wrote the manuscript.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability Data used in this study is available as supplementary data to this article.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethical approval This study received approval from the Human Research Ethics Committee of the University of Wollongong, Australia (Ethics number: 2021/116).

References

1. Ross J (2020) Coronavirus: almost all Australian universities head online. Times Higher Education. [Online]. Available https://www.timeshighereducation.com/news/coronavirus-almost-all-australian-universities-head-online. Accessed 11 Feb 2020.
2. Harwell D (2020) Cheating-detection companies made millions during the pandemic. Now students are fighting back. The Washington Post. [Online]. Available https://www.washingtonpost.com/technology/2020/11/12/test-monitoring-student-revolt/. Accessed 19 Feb 2021
3. Flaherty C (2020) Big proctor. Inside Higher Ed. [Online]. Available https://www.insidehighered.com/news/2020/05/11/online-proctoring-surging-during-covid-19. Accessed 19 Feb 2021
4. Munoz A, Mackay J (2019) An online testing design choice typology towards cheating threat minimization. J Univ Teach Learn Pract 16(3):54–70
5. Dendir S, Stockton Maxwell R (2020) Cheating in online courses: evidence from online proctoring. Comput Hum Behav Rep. https://doi.org/10.1016/j.chbr.2020.100033
6. Harmon OR, Lambrinos J (2008) are online exams an invitation to cheat? J Econ Educ 39(2):116–125
7. Heijne-Penninga M, Kuks JBM, Schoenrock-Adema J, Snijders TAB, Cohen-Schotanus J (2006) Open-book tests to complement assessment-programmes: analysis of open and closed-book tests. Adv Health Sci Educ Theory Pract 13:263–273
8. Hollister KK, Berenson ML (2009) Proctored versus unproctored online exams: studying the impact of exam environment on student performance. Decis Sci J Innov Educ 7:271–294
9. Vazquez JJ, Chiang EP, Sarmiento-Barbieri I (2021) Can we stay one step ahead of cheaters? A field experiment in proctoring online open book exams. J Behav Exp Econ. https://doi.org/10.1016/j.socec.2020.101653
10. Hubbard JK, Potts MA, Couch BA (2017) How question types reveal student thinking: an experimental comparison of multiple-true-false and free-response formats. CBE Life Sci Educ. https://doi.org/10.1187/cbe.16-12-0339
11. Hurtado S, Eagan MK, Pryor JH, Whang H, Tran S (2012) Undergraduate teaching faculty: the 2010–2011 HERI Faculty Survey. Higher Education Research Institute, UCLA, Los Angeles
12. Goubeaud K (2010) How is science learning assessed at the post-secondary level? assessment and grading practices in college biology, chemistry and physics. J Sci Educ Technol 19:237–245
13. Frisbie DA, Sweeney DC (1982) The relative merits of multiple true–false achievement tests. J Educ Meas 19:29–35
14. Dixon RA (1994) Evaluating and improving multiple choice papers: true–false questions in public health medicine. Med Educ 28:400–408
15. Couch BA, Hubbard JK, Brassil CE (2018) Multiple–true–false questions reveal the limits of the multiple–choice format for detecting students with incomplete understandings. Bioscience 68(6):455–463
16. Brassil CE, Couch BA (2019) Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. IJ STEM Educ. https://doi.org/10.1186/s40594-019-0169-0
17. Lesage E, Valcke M, Sabbe E (2013) Scoring methods for multiple choice assessment in higher education—is it still a matter of number right scoring or negative marking? Stud Educ Eval 39(3):188–193
18. Brabec JA, Pan SC, Bjork EL, Bjork RA (2020) True–false testing on trial: guilty as charged or falsely accused? Educ Psychol Rev. https://doi.org/10.1007/s10648-020-09546-w
19. Thompson J, Houston D (2020) Programmatic assessment condensed: introducing progress testing approaches to a single semester paramedic subject. J Univ Teach Learn Pract 17(3):14