A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus)

Dagogo J. Pepple,1 Lauriann E. Young,1 and Robert G. Carroll2
1Department of Basic Medical Sciences (Physiology Section), The University of the West Indies, Mona Campus, Kingston, Jamaica; and 2Department of Physiology, Brody School of Medicine, East Carolina University, Greenville, North Carolina

Submitted 24 September 2009; accepted in final form 20 January 2010

How We Teach

Pepple DJ, Young LE, Carroll RG. A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). Adv Physiol Educ 34: 86–89, 2010; doi: 10.1152/advan.00087.2009.—This retrospective study compared the performance of preclinical medical students in the multiple-choice question (MCQ) and long essay question components of a comprehensive physiology final examination. During the 3 yr analyzed, 307 students had an average score of 47% (SD 9.9) in the long essay questions and 64% (SD 9.9) in the MCQs. Regression analysis showed a significant correlation (r = 0.62, P < 0.01) between MCQs and long essay questions. When student performance was grouped by final course grade, a statistically significant correlation between MCQ and long essay scores existed only for the 210 students who received a passing grade (r = 0.20, P < 0.01). The MCQ and long essay question scores were not correlated for the 57 students who failed (r = 0.25, P = 0.06) or for the 40 students who achieved honors and distinctions (r = −0.27, P = 0.11). MCQ scores were significantly higher (P < 0.01) than essay scores for each of the groups when analyzed by two-way ANOVA. The results of this study suggest that for most students, the strong correlation between scores on MCQs and essay questions indicates that student performance was independent of testing format. For students at either end on the performance spectrum, the lack of correlation suggests that the performance in one of the testing formats had a strong influence on the final course grade. In addition, those students who failed the course were likely to be weak in both testing modalities, whereas students in all grade groups were more likely to perform better in the MCQs than in the long essay questions.

ASSESSMENT OF LEARNING is often one of the more difficult and time-consuming aspects of education. In physiology instruction in the preclinical years of medical school, knowledge is usually what is assessed. Course grades are a form of summative assessment, usually as a result of performance on examinations. Instructors have a variety of examination formats to use for student assessment, each with unique characteristics.

Multiple-choice questions (MCQs) are a common type of assessment due to their reliability, validity, and ease of scoring (5). MCQ tests are often constructed to assess the students’ ability to recall isolated pieces of information rapidly. There have been some suggestions, however, that competent students may perform poorly in MCQs because of their ability to read more into the questions than the examiners intended (1).

Another common type of assessment is open-ended, long essay questions. This format allows students more flexibility in their response and reflects their individuality of approach in which interpretative skills can be evaluated (6, 15). Essay questions also allow specific feedback to direct future learning (7). Nnodim (11) reported that MCQ papers were less sensitive predictors of the aggregate performance than essay questions, whereas Day et al. (4) observed that essay questions, although valid, failed to measure aspects of competence over and above those measured by MCQs.

The use of multiple formats for a single student assessment is labor intensive (3, 5). The additional labor may be justified if the formats test unique aspects of learning (14). If so, student scores in one format may not be tightly correlated with student scores in a different format. Some studies have reported statistically significant correlations between MCQ and essay scores in physiology continuous assessment tests (13) and short essay questions (10).

Before the implementation of the system-based curriculum at the University of the West Indies (UWI; Mona Campus), the comprehensive final examination in physiology had both essay and MCQ components. Because each component was weighted equally in calculating the final course grade, this provided an opportunity to compare student performance on the two testing formats. We analyzed historical data to determine if there was a correlation between the performance of our failing, passing, honors, and distinction students in the MCQs and long essay questions in the comprehensive final examination in physiology.

MATERIALS AND METHODS

This is a retrospective study involving 307 students who took a comprehensive final examination in physiology in 1997, 1998, and 1999. The data were collected from files in the Department of Basic Medical Sciences. There were 111 students in 1997, 94 students in 1998, and 102 students in 1999 who took the physiology examinations. The study was approved by the Ethics Committee of the Faculty of Medical Sciences of UWI (Mona Campus). Confidentiality was maintained as there was no disclosure of the names of the students from whom the data were derived.

Institutional setting. UWI is an international institution serving the Commonwealth countries of the Caribbean. It began as the University College of London in 1948 and achieved full university status in 1962 with authority to grant its own degrees. The Faculty of Medicine at the Mona Campus, which began in 1948 (now named the Faculty of Medical Sciences), was modeled on the British system and awards the Bachelor of Medicine and Bachelor of Surgery degree after 5 yr of

Address for reprint requests and other correspondence: D. J. Pepple, Dept. of Basic Medical Sciences (Physiology Section), The Univ. of the West Indies, Mona Campus, Kingston KgN 7, Jamaica (e-mail: dagogo.pepple@uwimona.edu.jm).
Table 1. MCQ and essay scores for students who failed, passed, and achieved honors/distinctions in the second Bachelor of Medicine and Bachelor of Surgery physiology examination

<table>
<thead>
<tr>
<th>Score Category</th>
<th>n</th>
<th>MCQ Scores, %</th>
<th>Essay Scores, %</th>
<th>MCQ Versus Essay Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed</td>
<td>57</td>
<td>50.49 (7.26)*</td>
<td>38.49 (7.34)</td>
<td>Correlation: 0.25</td>
</tr>
<tr>
<td>Passed</td>
<td>210</td>
<td>64.95 (6.43)*</td>
<td>48.32 (7.15)</td>
<td>P value: 0.06</td>
</tr>
<tr>
<td>Achieved honors/distinctions</td>
<td>40</td>
<td>76.31 (6.00)*</td>
<td>59.32 (4.19)</td>
<td>Correlation: -0.27</td>
</tr>
<tr>
<td>Overall</td>
<td>307</td>
<td>63.77 (9.87)*</td>
<td>47.33 (9.89)</td>
<td>P value: 0.62</td>
</tr>
</tbody>
</table>

Values are means (SD); n, numbers of students. *Multiple-choice question (MCQ) scores were significantly higher than essay scores by two-way ANOVA followed by Tukey’s test.

study. Course grades are also modeled on the British system, where <50% is a failing grade, 50–64% is a passing grade, 65–70% is an honors grade, and >75% is a distinction grade. This grading system is very different from the standard 10-point scale used in much of the United States, where <70% is a failing grade, 70–79% is a grade of “C,” 80–89% is a grade of “B,” and >90% is a grade of “A.”

Examination formats. Before the implementation of the systems-based curriculum in 2001, the final examination in physiology was taken at the end of the 2-yr preclinical training. It had both essay and MCQ components.

The essay component of the final examination lasted 2.5 h. Students were presented with six open-ended questions, where they selected five questions and had 30 min to write on each of them. The MCQ component of the final examination consisted of 150 questions with a duration of 3 h. The MCQ paper was divided into three sections. The main section consisted of questions with one correct answer out of four options. Two smaller sections included questions with one or more answers out of four options and a statement and reasoning area. Both the essay and MCQ modalities had equal weighting and contributed 75% to the final score for each student. The remaining 25% of the final score for each student was the average of three in-course examinations. A viva voce (oral) examination was given only to those students who were on the border line of pass/fail (i.e., those achieving a score between 45% and 49%) and those who were being considered for honors or distinction (i.e., those achieving scores between 60% and 64% and between 70% and 74%, respectively).

Data analysis. The data were analyzed with the SPSS statistical package and are reported as means (SD). A Kruskal-Wallis test was used to identify any variation in scores between classes over the 3 yr of data collection to determine if the aggregate student performance could be combined to provide a larger sample size for analysis. Pearson’s correlation coefficient (r) was calculated to determine if performance on the essay question was correlated to the performance on the MCQ score for each group. Finally, two-way ANOVA followed by Tukey’s test was used to determine both the impact of examination format (MCQ vs. essay) and the final course grade (fail, pass, honors/distinction) on student performance. The level of significance was set at P < 0.05. During the period of this study, item analysis was not performed on the MCQ examinations at UWI.

RESULTS

Analysis of the data revealed no statistically significant differences in student performance in the 3 yr studied; hence, the respective MCQ and essay scores were combined and analyzed for the 3 yr under review. Because of the small number of students who had distinctions, their scores were grouped together with those who had achieved honors.

Regression analysis showed a significant correlation (r = 0.62, P < 0.01) between MCQs and long essay questions. When students were grouped by final course grade, a significant correlation between MCQ and long essay scores existed only for the 210 students who received a grade of passing (r = 0.20, P < 0.01). The MCQ and long essay question scores were not correlated for the 57 students who failed (r = 0.25, P = 0.06) or for the 40 students who achieved honors and distinctions (r = −0.27, P = 0.11).

MCQ scores were significantly higher (P < 0.01) than essay scores in all grade groupings when analyzed by two-way ANOVA followed by Tukey’s test. The detailed data are shown in Table 1.

Figure 1 shows scatterplots of the aggregate student performance. As indicated in Table 1, there was a significant correlation between essay and MCQ scores for the entire group. MCQ performance was stronger than essay performance for all but five of the students.

Figures 2–4 show examination performance broken down by course grade. For many of the students receiving a final passing grade (Fig. 2), the stronger MCQ performance compensated for a failing grade in the essay component. This pattern was repeated for students receiving honors/distinctions (Fig. 3). For students who failed the examination, the poor performance on the essay component was strong enough to offset a passing grade on the MCQ component (Fig. 4). Only one student who ultimately failed the examination had a passing score on the essay component.
DISCUSSION

The most important finding of this study was that there was a statistically significant overall correlation between student performance on MCQ and essay examination scores. The correlation between long essays and MCQ scores indicated that, in general, students who performed well in the essays were also likely to do well in the MCQs. Oyebola et al. (13) and Moqattash et al. (10) noted a similar relationship in the performance of their students.

We also observed that scores for MCQs compared with long essay questions were consistently higher in all groups, that is, students who had failed, passed, or received honors/distinctions. The average difference in the scores was 12 points for the group of students who failed the overall examination and was even larger for the other groups. It is worth mentioning that the correlation between long essay scores and MCQ scores for students who failed may not have achieved statistical significance ($P = 0.06$), partly due to the small sample size ($n = 57$) of this subgroup. A future study with a larger sample size might help to clarify this doubt. However, the effect of a small sample size may not be responsible for the lack of correlation observed in the students who achieved honors and distinctions, even though they had a similar sample size ($n = 40$). This is due to the nonsignificant $P$ value of 0.11 in this subgroup.

As schools modify their assessment approach, this type of difference in student performance can have a significant impact on the final grades earned in a course. This difference is likely not limited to this student population, as a lower performance in the physiology long essay questions, particularly when the question lacked cues, was also noted by McCloskey and Holland (9).

An interesting finding is that there was no correlation between long essay and MCQ scores for those students who performed at honors and distinction levels. We interpret this to mean that the more competent students had unique strengths in either one or the other examination format. For example, students with strong factual recall abilities may have scored higher in the MCQ component, whereas students who had strong analytical or interpretative skills, and an ability to organize and apply knowledge, were able to score higher in the essay component of the examinations. Anbar (1) also observed a lack of correlation in the performance of competent students in MCQ tests compared with open-ended tests and a positive correlation for the less competent students.

The consistently lower scores in the long essay questions may be attributed to the bias or more subjective marking schemes of different lecturers compared with the more quantitative nature of MCQs. A similar observation of higher scores in MCQs and short essay questions compared with structured, integrated long essays was reported by Moqattash et al. (10), who indicated that one of the reasons for the higher scores was the elimination of examiners' bias. However, we believe that long essay-type assessment is a sensitive test requiring students not only to recall facts but also to use higher-order cognitive skills, such as analytical, interpretive, and application skills.
The scores on the MCQ component were much lower than would be expected for examinations in the United States system using a 10-point grading scale. The minimum performance for passing in the British grading system is 50%, in contrast to the 70% minimum passing grade for United States schools. Despite the different scales, internal and external review of curriculum and student performance have emphasized the validity of the grading schemes using the standard setting (2, 12), particularly when determining the pass/fail threshold. An interesting exception to the 70% passing grade in the United States is the licensing examination of the National Board of Medical Examiners, where the pass/fail threshold is often close to 56% of items answered correctly (8).

In summary, this retrospective study indicates that, in general, there was a strong correlation between student performance in MCQs and long essay questions. This suggests that the additional effort used to prepare examinations in both formats may not be justified. There was no correlation between MCQ and long essay scores for either the more competent students or the students who received failing grades. Thus, student grades determined by an examination format that includes both testing modalities may be different than the grades obtained by using only one of the modalities. The overall higher performance in all groups of students on the MCQ examination indicates that without the long essay examination format, some of the failing students may likely have passed the final examination in physiology.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES