Scientific Research

Short Answer Questions or Modified Essay Questions— More than a Technical Issue

Sven Wallerstedt[1], Gudrun Erickson[2], Susanna M. Wallerstedt[3]

[1]Department of Emergency and Cardiovascular Medicine, University of Gothenburg, Gothenburg, Sweden; [2]Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden; [3]Department of Clinical Pharmacology, University of Gothenburg, Gothenburg, Sweden.
Email: susanna.wallerstedt@pharm.gu.se

ABSTRACT

Purpose, The present article was built on the assumption that the form of an examination may influence learning, and may also reflect different kinds of knowledge. The aim of the study was to evaluate whether the results of an examination differ when short answer questions (SAQ) or modified essay questions (MEQ) are used. Method, Forty-nine students in the internal medicine course in Gothenburg, Sweden, performed a written examination in 2003, which included both SAQ and MEQ. Result, The correlation between the results of SAQ and MEQ was 0.59 (P < 0.001). The percentage correctly answered questions in the two types did not differ significantly. Some students had poor results in either SAQ or MEQ. Conclusion, The general outcome of the study indicates that results of SAQ and MEQ demonstrate a significant correlation. However, they may also reflect differences in mastery of the knowledge domain, which should be considered in relation to aspects of validity.

Keywords: Evaluation; Examination; Medical School; Modified Essay Questions; Validity

1. Introduction

Examinations have several functions, e.g. to make sure that students have learnt the essential part of a course, and to give feedback to students and teachers on the effectiveness of learning and teaching. It may also be assumed that the form of examination influences the pedagogical process at large, thereby contributing to what is sometimes referred to as systemic validity [1]. Thus, summative as well as formative aspects of assessment should be recognized. Moreover, examinations need to be analyzed within the framework of an expanded view of validity, in which coverage of the defined knowledge domain—the construct—as well as inferences, actions and consequences are taken into account [2].

The questions in a written examination can be constructed in different ways, e.g. short answer questions (SAQ) or essay questions [3]. Within the medical field, the latter type has been developed into modified essay questions (MEQ) in order to assess clinical problem-solving skills [4]. A clinical case is presented in a chronological sequence of items in a case booklet. After each item a decision is required, and the student is not allowed to preview the subsequent item until the decision has been made. The reliability of MEQ is reported to be reasonable [5]. The aim of the present study was to evaluate

if, and how, the results of a student examination differ when SAQ and MEQ are used.

2. Method

The study was carried out in the 2003 internal medicine course in Gothenburg, Sweden. At the end of the course, the students take a written examination in order to pass the theoretical part of the course. 49 students took part in the present examination, which consisted of 15 SAQ and 5 MEQ. All questions had previously been constructed, used, and validated by other universities in Sweden, which, at that time followed the same curriculum. The question constructors had not been informed that their questions were to be used in order to compare results on SAQ and MEQ. None of the questions had been available for the Gothenburg students. Based on the empirically developed guidelines, rating was done by one of the authors (SW), with the possibility of co-rating if needed.

Statistics

Statistical analyses and randomization were conducted using SPSS 12.0. Spearman correlation coefficients were calculated between the student results from 1) SAQ and MEQ; 2) SAQ and two randomly chosen MEQ; 3) SAQ

and the remaining three MEQ, and 4) the MEQ in 2) and 3). Percentage correctly answered questions in different groups were compared with Mann-Whitney's test. Values are presented as mean (standard deviation).

3. Results

The examination results for the 49 medical students are presented in **Table 1**. The percentage correctly answered questions did not differ significantly between SAQ and MEQ (P = 0.075).

Two students' results were markedly poorer than the rest. Their percentages correctly answered questions were 33% and 51%, respectively (SAQ 60% and 60%, respectively; MEQ 22% and 47%, respectively). For SAQ, two students had less than 50% correctly answered questions (43% each), their results in MEQ being 69% and 64%, respectively.

There was a correlation between the results of 1) SAQ and MEQ; 2) SAQ and two randomly chosen MEQ; 3) SAQ and the remaining three MEQ; and 4) MEQ in 2) and 3) (**Table 2**).

4. Discussion

Although a few individual students had poor results in either SAQ or MEQ, their results in one of the parts being markedly better, the present study generally showed a significant correlation between SAQ and MEQ. This may not be surprising, since it has been reported that a large proportion of MEQ are often factual recall questions or

Table 1. Various test models with maximum score for each type, and results (% correct answers) for all 49 medical students participating in the examination (SAQ, short answer questions; MEQ, modified essay questions).

	Maximum score	Results, mean (SD)
SAQ	20	72 (14)
MEQ	48	67 (11)
Two randomly chosen MEQ	22.5	65 (13)
The remaining three MEQ	25.5	69 (12)

Table 2. Spearman correlation coefficient between the student results of short answer questions (SAQ) and modified essay questions (MEQ).

	Correlation coefficien	P
SAQ vs MEQ	0.58	<0.001
SAQ vs two randomly chosen MEQ	0.46	0.001
SAQ vs the remaining three MEQ	0.48	0.001
The two randomly chosen MEQ vs the remaining three MEQ	0.32	0.025

deal with interpretation of data [6], *i.e.* they imitate traditional SAQ. Although the MEQ format was initially developed to assess clinical problem-solving skills, Feletti and Smith [6] found that only 27% of MEQ on nine occasions during three consecutive years were actually problem-solving questions. The correlation coefficients between SAQ and MEQ found in our study were similar to the ones between subgroups of MEQ. This indicates that variation in content rather than in format affects the outcome.

The results of the present study do not allow conclusions concerning what kind of knowledge was assessed through the two types of questions. However, the differing results for the low achieving students may indicate that different types of knowing are focused upon. Indeed, MEQ has been claimed to be especially valuable for diagnosing student weaknesses [5].

To further explore the issue of different question types, larger number of results would be required. However, since 2003, the faculties of medicine in Sweden have developed diverging curricula. Thus, it was not possible to further validate our results by expanding our study to include more students. Moreover, it would be valuable to actively collaborate with the students, for example by using established reporting techniques during test taking, interviewing students about their perceptions after the exam, or asking them to write comments on specific aspects of the different formats used, e.g. in relation to degree of coverage of the domain, and the perceived authenticity of the examination.

A disadvantage with MEQ is that the examination time usually only allows a small number of cases, thus limiting the number of items to test. This may negatively affect the coverage of the area (content validity), as well as the precision of the measurement (reliability). Another problem with MEQ is that the construction of such items is time-consuming. However, it may also be claimed that there are validity related advantages to the technique, for example concerning the aspect of authenticity, commonly regarded as an important feature of construct validity.

5. Conclusion

Our experience is that students appreciate the MEQ type of questions. However, one may ask if it is worth the effort, since the results correlate well with the much more easy-to-construct SAQ. Maybe a written examination consisting of both SAQ and MEQ, like in the present study, could be an adequate compromise, with beneficial effects on validity and reliability, as well as on feasibility.

REFERENCES

[1]   J. R. Frederiksen and A. Collins, "A Systems Approach to

Educational Testing," *Educational Researcher*, Vol. 18, No. 9, 1989, pp. 27-32.

[2] S. A. Messick, "Validity," In: R. L. Linn, Ed., *Educatio-nal Measurement*, 3rd Edition, American Council on Edu-cation/Macmillan, New York, 1989, pp. 13-103.

[3] G. K. Cunningham, "Assessment in the Classroom: Con-structing and Interpreting Tests," Falmer Press, London, 1998.

[4] G. I. Feletti and C. E. Engel, "The Modified Essay Ques-tion for Testing Problem-Solving Skills," *The Medical Journal of Australia*, Vol. 1, No. 2, 1980, pp. 79-80.

[5] G. I. Feletti, "Reliability and Validity Studies on Modi-fied Essay Questions," *Journal of Medical Education*, Vol. 55, No. 11, 1980, pp. 933-941.

[6] G. I. Feletti and E. K. Smith, "Modified Essay Questions: Are They Worth the Effort?" *Medical Education*, Vol. 20, No. 2, 1986, pp. 126-132. doi:10.1111/j.1365-2923.1986.tb01059.x