Multiple Choice Questions for Continuing Education in Family Medicine

Gordon L. Dickie, MB, ChB, Martin J. Bass, MD, MSc, Walter O. Spitzer, MD, MPH, and Robin S. Roberts, MTech
London, Ontario, Montreal, Quebec, and Hamilton, Ontario

Multiple choice questions used in continuing medical education may require characteristics different from those used in tests and examinations. The questions of three continuing education exercises were assessed by 48 board certified family physicians. Each physician answered one third of the questions and then judged them on usefulness in patient management, as discriminators of quality of medical care, and on educational value. The relevance of each question was determined by a composite index based on these factors. The content of each question was independently analyzed by three physicians.

Relevant questions tended to be those concerned with surgery, symptoms, and management, and those requiring more than simple factual recall. Questions on office management or concerning specific diagnoses were considered less relevant. Neither the manner of asking the question nor the severity of the illness appeared to influence relevance. This information should benefit those developing continuing education programs for family physicians, especially those designed for self-administered individual learning.

Throughout the last decade there has been an increasing awareness of the need for continuing education for practitioners of all branches of medicine. Nowhere has this need been felt more acutely than in the field of family medicine. Unlike his colleagues in other specialties, the family physician often works in relative professional isolation and has fewer external assessments of his professional performance. Many types of continuing medical education have been instituted to help the family physician and have enjoyed a greater or lesser success as measured by the subjective impressions of participants. One such mode that enjoys considerable popularity is the self-assessment exercise using multiple choice questions. These have been developed and distributed by academies and

From the Department of Family Medicine, University of Western Ontario, London, Ontario, the Departments of Family Medicine and Epidemiology, McGill University, Montreal, Quebec, and the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario. Requests for reprints should be addressed to Dr. Gordon L. Dickie, St. Joseph's Family Medical Centre, 362 Oxford Street East, London, Ontario N6A 1V8.

0094-3509/81/131031-05$01.25

colleges of family medicine. The purpose of this paper is to examine the content of one such program to determine the characteristics of those questions which seem to be most relevant to family physicians.

Methods

The questions examined in this study were prepared by the Connecticut and Ohio Academies of Family Physicians and constitute part of their Core Content Review Program for 1974-75. The initial part of the study has already been reported.[1] A panel of 48 board certified family physicians, 24 from academic practice and 24 from private practice in the same communities, assessed the Core Content Review. Three of the six exercises offered in 1974 and 1975 were examined. One exercise was randomly allocated to one third of each group so that each of 308 questions was assessed by eight academic and eight community family physicians. In addition to answering the questions, the physicians answered probes concerning the subject matter of each question. These dealt with the frequency of occurrence of the problem, its ability to discriminate between family physicians providing high and low quality care, the usefulness of the question as a learning experience, and its usefulness in the management of patients.

Several composite indices of "relevance" incorporating these variables, and also correctness of response, were derived. After much deliberation the investigators adopted the following definition: A question is relevant if 12 or more of the 16 physicians answering it indicated that it is thought to discriminate between the family physician providing good quality care and the physician providing less than adequate care, *OR* that the question deals with a problem reported to be seen at least once a year, and it is useful either for physician education or in patient management.

A composite criterion of relevance was selected by the authors to avoid the wide variation normally associated with subjective assessments of relevance. According to this predetermined standard, quality discrimination by itself, as judged by the responding family physicians, was considered a sufficient index of relevance. The authors felt that peer group assessment of this variable was a powerful determinant of relevance, but that other variables should also contribute to the definition. The concepts of educational and management usefulness were also felt to be important. It was recognized that many questions might be educationally useful, but unless they dealt with problems reportedly seen at least once a year, they could not be regarded as relevant to the continuing medical education of family physicians. Thus, the standard adopted permitted inclusion of questions containing clinically useful information on the conditions commonly seen in family practice.

The content of the questions was independently analyzed by three physicians. Each question was examined from several points of view. The content matter was first classified according to the medical specialty involved. Next, the content was categorized according to whether it involved dealing with a complaint or symptom, a specific diagnosis, clinical investigation, management, or aspects of the basic medical sciences. The subcategories were not designed to be mutually exclusive, and questions might be assigned to more than one. Last, an

Table 1. Assessment of 308 Core Content Questions

Perceived Attributes	Percentage of Questions Designated by 12 to 16 Physicians
Useful as learning experience	72
Relevance*	54
Quality discriminator	44
Problem occurs more than once per year	34
Useful for patient management	23

*For composite definition, see section on "Methods"

assessment was made of the seriousness of the condition.

The structure of the question was next examined and classified according to whether the correct answer might be suggested by the length or completeness of one of the given responses, whether the question was asked negatively, and whether more than one of the given responses was correct. The official answers to each question were also examined, and it was determined whether an adequate explanation of the official answer was given and whether all alternatives were dealt with.

The third classification of the questions was in respect to the intellectual process involved in answering them. Using the method described by McGuire,[2] there was little variation between the three physicians' assessments, but where there was disagreement, two of the present authors (M.J.B., G.L.D.) independently performed an additional assessment, and in every case reached agreement on classification.

Statistical analysis was undertaken using Chi-square tests with Yates correction. Although values of P less than .05 are conventionally considered to be statistically significant, in our analysis, where Chi-square was calculated for multiple factors, P<.05 as a critical value may be too liberal, and the data should be interpreted accordingly. Logistic function analysis was also undertaken to examine the independent contribution to relevance of the factors studied.

Table 2. Relevance of Question Content Areas

Content	Number	Percent Relevant
Specialty orientation		
Adult medicine	168	53
Pediatrics	59	53
Psychosocial (includes family dynamics)	33	52
Obstetrics	22	73
Surgery (includes gynecology)	27	78*
Strategies of practice		
Prevention (includes health maintenance)	1	100
Office organization	11	9**
Clinical activity		
Related to given diagnosis	67	39**
Complaint or symptom	36	72*
Clinical investigation	48	60
Basic science	54	46
Management/therapy	97	68**
Severity of illness		
Conditions with minor degrees of severity	154	60*

*P<.05
**P<.01

Results

Table 1 shows the distribution of specific attributes of the 308 questions. Of the 72 percent considered educationally useful, those relating to adult medicine were significantly more useful, whereas those relating to psychosocial aspects were significantly less educationally useful. Of the 44 percent of questions considered to be discriminators of high or low quality clinical practice, questions on surgery and patient management were significantly more frequent. Only 23 percent of the questions were considered useful in patient management, with no particular content area being specifically identified.

Of the 308 questions 54 percent achieved the composite criterion of relevance. Table 2 shows the number of questions with given content areas and the percentage of relevant questions in each group. These data indicate that questions involving management and therapy have high relevancy rates, whereas questions on office management and specific diagnoses are of little relevance. In addition, items involving surgery or symptoms may have higher relevancy rates, but the strength of the evidence is somewhat weaker.

Also examined was question structure and answer content (Table 3). The physician reviewers considered that in 94 percent of cases an adequate explanation accompanied the given answer. In

Table 3. Question Structure, Answer Content, and Relevance

Content	Total Number of Questions	Number Relevant	Percent Relevant
Adequate examination of answer	289	162	56
All alternatives dealt with in answer	157	75	48
More than one correct alternative	58	28	48
Question asked negatively	13	6	46
Correct answer suggested by length	7	6	86

only seven questions was the answer suggested by the length or complexity of the alternatives.

The results of the examination of the different intellectual processes involved in completing the Core Content Review are shown in Table 4. Nearly one half of the questions required only simple recall of isolated information to answer them. Questions considered most relevant, however, require recognition of meaning, total situation evaluation, or problem solving of a familiar type.

The analyses described here so far have considered the presence or absence of individual features of content structure and intellectual process. It is possible that such individual factors act together to contribute to the relevance of questions. A logistic function analysis of the data was undertaken to examine the contribution to relevance or nonrelevance of each of the various characteristics of questions taken individually while controlling statistically for the effects of the others. This analysis confirmed many of the findings in the initial analysis; in particular, factors shown to be positively associated with relevance were the content areas of surgery, psychosocial problems, and questions in which the given answer was fully explained. Factors negatively associated with the relevance were questions regarding a specific diagnosis and those dealing with the basic sciences. These findings appear to be consistent with the other individual analyses reported here.

Discussion

The apparently low overall relevance figure of 54 percent may seem suprising in a program such as this, designed specifically for family physicians, but the criterion of relevance was stringent, a fact that probably results in a conservative estimate. In formulating the criterion, the authors recognized that many questions might be educationally useful, but unless they dealt with problems reportedly seen at least once a year, they could not be regarded as relevant to the continuing medical education of family physicians. It was considered important to avoid the inclusion of questions that would educate on conditions rarely seen in family practice.

There is a tendency in the difficult and complex issues occurring in medicine for the longest alternative response to a multiple choice question to be the correct answer, thus giving a clue to the respondent. Although this feature decreases the ability of the question to differentiate the knowledgeable from the ignorant, it may enhance learning by directing the physician to the correct answer while challenging him or her with other considerations. It is intriguing that six of the seven questions of this nature in the study were deemed relevant. This is one of the areas in which questions for continuing medical education may properly differ from those used in examinations.

Table 4. Intellectual Process and Relevance

Process	Number	Percent Relevant
Recall of isolated information	143	51
Recognition of meaning or implication	44	59
Relevant generalization to explain phenomenon	17	29
Simple interpretation	10	60
Application of simple principles in a familiar situation	76	59
Application of combination of principles in new situation	7	14
Evaluation of total situation	11	82
Total	308	

Note: This classification is based on that described by McGuire[2]

Classifying the intellectual processes used to answer a question was sometimes difficult. The McGuire classification was designed to reflect increasingly complex thinking processes from stage 1 (recall) to stage 7 (situation evaluation). For some questions it was difficult for the two experienced family physician reviewers to decide whether they were using recall (stage 1) or applying principles (stage 5). Further work is required to categorize continuing education multiple choice questions. Questions requiring more than simple recall had increased rates of relevance, and this may pinpoint an important category of questions which should be taken into account by those designing such exercises.

Although the respondents were asked to assess the usefulness of the questions for education, they were not subjected to a retest to determine whether any learning had taken place. This may be an appropriate question for further study.

That questions on office management are not considered relevant may relate to the extensive professional experience of the panel, who had all been in practice for five or more years. Perhaps more junior physicians would find such questions useful.

Self-administered multiple choice questions constitute part of the maintenance of certification process recently introduced by the College of Family Physicians of Canada[3] and play an important part in the continuing medical education of many family physicians. The characteristics of the best questions for this purpose must be defined so that the physicians involved will obtain the maximum benefit from participation. It would seem that questions thought to discriminate between high and low quality clinical care are most important for this purpose, but there is a need to demonstrate a correlation between answering such questions and providing good quality care in practice. Confirmation of the external validity of these findings would be of great assistance to those who design multiple choice questions for continuing medical education.

Acknowledgements

The authors thank Dr. Allan Donner for his statistical advice and helpful suggestions. We also acknowledge the assistance and cooperation of the Board of Directors and senior staff of the Core Content Review of Family Medicine, and the financial support of the Physicians' Services Incorporated Foundation, Toronto.

References

1. Spitzer WO, Dickie GL, Bass MJ, et al: The relevance to family physicians of core content review: Evaluation of a program of continuing education. Can Med Assoc J 122: 426, 1980

2. McGuire C: A process approach to the construction and analysis of medical examinations. J Med Educ 38:556, 1973

3. Rice DI: Maintenance of certification. Can Fam Physician 25:979, 1979