An Assessment of the Influence of Cueing Items in Objective Examinations

WILLIAM H. McCARTHY, M.B.B.S., F.R.A.C.S.*
University of Illinois College of Medicine, Chicago

Most objective examinations allow the examinee to select a correct response from a list of possible alternatives. This format assumes that the examinee can recall unaided the response he recognizes as correct among the alternatives. Recognition with the aid of visual cues and unaided recall may be 2 separate abilities not necessarily coexistent in every situation.

This study was designed to determine whether differences in achievement could be identified when medical students were tested on simulated clinical problems presented with and without visual cues. It was hypothesized that the opportunity to select a response from a list of possibilities would enable many students to identify the correct response in situations where the appropriate information could not be remembered.

METHOD

Twenty third-year medical students volunteered for the study. Performance by these students on the comprehensive examinations of the University of Illinois College of Medicine suggested that they were representative of the total class.

TASK

Alternate formats of 2 simulated clinical problems were constructed; one made use of visual cues and one did not. Each problem involved sequential analysis of a patient with upper abdominal pain. The problems were similar in historical information, physical examination, and diagnostic workup. Corresponding sections of Problem I and Problem II had the same number of relevant items but the primary diagnosis and therapeutic regimen differed; this allowed comparable scoring procedures. In Figure 1 the main features of the 2 problems are compared.

Each problem was presented as (*a*) a noncued, oral, free-response examination and (*b*) a cued, printed, erasure-type examination.

In the oral free-response format the problem was introduced by the examiner with a small amount of information such as might be volunteered by an actual patient. The examinee then depended entirely on his own resources to generate further requests for historical, physical, laboratory, diagnostic, and therapeutic data. The examiner avoided inadvertent cueing by reading aloud, from previously prepared index cards, all the requested data. For each problem the information on the data cards matched exactly that which was presented in printed form.

In the printed erasure examination visual cueing is provided by lists of possibly relevant cueing items in each section of the problem. Information is gained by erasing opaque overlays beside relevant items. The examinee proceeds through the sections of the problem erasing overlays until he has enough information to select an appropriate diagnosis and plan of management.

* Commonwealth Fund Fellow, Office of Research in Medical Education.

	Problem I	Problem II
History	Weight loss & abdominal pain	Weight loss & abdominal pain
	Symptoms of cardiac failure	Symptoms of cardiac failure
	Urinary symptoms	Urinary symptoms
	Painful left calf	Diarrhea
Physical	Abdominal mass	Abdominal mass
	Cardiac failure	Cardiac failure
	Enlarged prostate	Enlarged prostate
	Lymph node in neck	Steatorrhea
Laboratory	Same tests appropriate in both	
Diagnosis	Cancer of pancreas	Chronic pancreatitis
	Cardiac failure	Cardiac failure
	Myocardial Ischaemia	Myocardial Ischaemia
	Prostatic hypertrophy	Prostatic hypertrophy
	Phlebothrombosis	Gallstones
	Smokers cough	Smokers cough
Therapy	Digitalis	Digitalis
	Diuretics	Diuretics
	Cessation of smoking	Cessation of smoking
	Celiotomy	Cholecystectomy

FIGURE 1
Main features of 2 simulated clinical problems.

One-half of the experimental group took Problem I in the noncued oral format and Problem 2 in the cued printed format. The remaining 10 students took Problem II orally and Problem I in printed form. In all cases the oral examination preceded the printed examination.

Achievement of each student in the 2 test situations was compared. The following measures were obtained for each type of examination: (*a*) total number of selections in each section; (*b*) quality of the selections, that is, the number of correct responses in each section; (*c*) ranking differences, that is, correlation between total scores; (*d*) average improvement in the presence of visual cueing; (*e*) influence of the cued format on those students performing poorly in the oral examination; (*f*) relevance of the additional items erased in the laboratory section of the printed examination.

FINDINGS

Increase in the number of items selected in each section of the printed problem was obvious. In the presence of visual cues there was a substantial increase in both total and correct selections. Table 1 shows the means of the total and correct selections in each section for both examinations.

To assess ranking differences in the 2 formats Pearson product moment correlation between total scores was computed. A correlation coefficient of −0.09 indicated no significant correlation between the 2 formats.

In the cued situation an average improvement of 41.2 marks (25 per cent increase) occurred. The mean of the total score for the oral problem was 109.3; the corresponding mean for the printed problem was 152.6.

Raw scores suggested that those students who had the most difficulty in the oral free response situation were particularly aided by the printed format. To test this hypothesis the oral performance was ranked and the gain in

marks by each student recorded. The means of the gains of the 10 upper and the 10 lower students were computed. For the former the mean gain was 24.6± SD 42.04; for the latter it was 62.6± SD 22.33. This difference is significant at the 0.05 level. Thus, in fact, the objective format does aid particularly those students who had most difficulty in the oral noncued problem.

It is important to know if the examinee is encouraged to select mostly relevant responses or if the presence of cueing items also encourages the selection of harmful interventions. The selections in the laboratory section were studied to assess their relevance. Scores in this section improved considerably in the cued situation (average gain 74 marks) but the scores did not correlate significantly with the total selections in this section ($r = 0.16$). Thus, while the additional laboratory tests ordered were in general helpful, many examinees were encouraged to select possibly harmful procedures.

DISCUSSION

The effects of guessing on objective examinations have been extensively investigated (1-3) and corrective scoring procedures (4, 5) have been devised. The effect of visual cueing as an aid to the recall of relevant information is less clearly identified. Studies on the influence of this factor are in conflict (6, 7), perhaps indicating that different abilities have been examined in the various studies. Ebel (7) has stated what he believes is the consensus among educators: ". . . it is reasonable to conclude that items—short-answer and choice—yield meaningfully different kinds of information concerning the examinee's abilities."

The findings described here suggest important differences in achievement on a test of clinical competence using an uncued oral and a cued printed problem.

TABLE 1

TOTAL AND CORRECT SELECTIONS

	Mean of Total Selections*		Mean of Correct Selections*	
	Uncued (oral)	Cued (printed)	Uncued (oral)	Cued (printed)
History	12.8	22.1	12.2	16.2
Physical	11.1	15.9	10.05	12.0
Laboratory	8.1	19.0	6.4	13.5
Diagnosis	1.9	4.3	1.2	3.3
Therapy	1.4	3.6	1.1	2.2

* These mean differences are significant at the 0.01 level of confidence.

The correlation of −0.09 between total scores indicates a completely different rank order on the 2 formats; this suggests that the 2 formats may indeed be evaluating quite different abilities.

The cued examination appeared to aid particularly the students less able to recall pertinent information in the oral noncued situation. These students were able to recognize as relevant much information not recalled without aid. The objective format may be most in error in the evaluation of these students.

Clinical competence involves at least 3 kinds of ability. The physician elicits relevant data from the patient in an oral confrontative situation; he performs a thorough but discriminating laboratory workup; he synthesizes the obtained data into a diagnosis and plan of management. Data gathering is essentially dependent on uncued recall of information. The erasure-type problem, and perhaps cued objective examinations of all types, appear to contain an influence which makes more difficult the assessment of clinical competence by these formats. In this study almost all students performed significantly better when visual cues were provided.

Most of the findings of this study suggest that the objective format might be less suitable than other techniques for estimation of certain parameters of physician performance. History-taking and

physical examination are essentially noncued activities. This raises questions concerning the validity of cues in tests of these abilities. In this study the mean number of responses rose most sharply in the history and physical sections.

Diagnostic workup normally involves the use of many cues. The objective format might be more realistic in evaluation of this activity. Though the presence of cues in this section significantly improved performance (Table 1), this may not be so important. The physician normally uses the results of early investigation to cue the ordering of further diagnostic workup.

The printed erasure problem appears to constitute a significant advance in the evaluation of clinical competence. It enables the examinee to perform a total evaluation of a clinical problem and it permits this to happen in the sequential fashion utilized by the practicing physician. However, in this format visual cueing is maximized and further attention may have to be given to this factor before this format achieves its correct place in the total evaluation of clinical competence.

SUMMARY AND CONCLUSIONS

Achievement of 20 medical students in an uncued oral examination is compared with achievement on a similar cued printed examination. The abilities assessed by the 2 formats appeared to be quite different. Scores were in general higher when visual cues were provided. Cueing items seemed to aid particularly the students unable to recall relevant information.

The suggestion is made that the cued erasure-type problem, and perhaps other formats involving selection from a list of responses, may not provide optimal evaluation of certain aspects of physician performance. It is suggested that data gathering by a physician is essentially a noncued activity and this may be more adequately assessed by some type of noncued objective oral examination such as that detailed in this paper.

REFERENCES

1. SHERRIFFS, A. C., and BOOMER, D. S. Who is Penalized by Guessing? *J. Ed. Psychol.*, 45:81-90, 1951.
2. SWINFORD, R., and MILLER, P. Effects of Directions Regarding Guessing on Item Statistics of a Multiple Choice Vocabulary Test. *J. Ed. Psychol.*, 44: 129-139, 1953.
3. SWINFORD, F. Analysis of a Personality Trait. *J. Ed. Psychol.*, 32:438-444, 1941.
4. Committee on Student Appraisal, University of Illinois College of Medicine. Setting Standards of Competence—The Minimum Passing Level. January, 1964 (mimeographed).
5. LINDQUIST, E. F. (Ed.). *Educational Measurement.* Washington, D.C.: American Council on Education, 1961, Pp. 347-351.
6. HURLBURT, D. The Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meaning. *J. Ed. Res.*, 47:561-576, 1954.
7. EBEL, R. L., and DAMRIN, D. E. Tests and Examinations. In *Encyclopedia of Educational Research.* New York: Macmillan Company, 1960, Pp. 1510-1511.

Discussion

DR. RUTH PEACHEY (Hahnemann): Do you have any information as to whether or not student anxiety might have been a contaminating factor?

DR. MCCARTHY: I have no objective evidence about it. The examination was taken by a group of third-year clerks whom I and others taught. I had, I think, achieved a certain amount of rapport with these students and I am not on the evaluation team of the University of Illinois.

As an outsider, I was perhaps able to get closer to them than an inside person might have. I explained to them that the examination wasn't for evaluation purposes, that it was part teaching and part research for me. I do not think anxiety was much of a factor, but I cannot exclude it.