

Testing in English Literature: Toward a Better Rationale

Muriel W. Tomkins

In September 1955, Edward J. Gordon presented in the *English Journal*, five useful "levels of teaching and testing" in English literature and thus helped to point the way toward a better rationale. His levels suggest a hierarchy of questions that place increasing demands on the student's talent for thinking in abstract terms. The abilities which they test are as follows:

(1) to remember a fact, (2) to prove a generalization that someone else has made, (3) to make one's own generalization, (4) to generalize from the book to its application in life, and, finally, (5) to carry over the generalization into one's own behavior.¹

Level 1 questions test only the student's memory, his ability to recall characters and events in a literary work he has studied, or to identify lines quoted from it. Level 2 questions require a degree of independent thinking, as this sample suggests: "If you were to judge by *Babbitt*, what attitudes are held by the 'typical' businessman?" Level 3 questions demand that the student arrive at his own interpretation — to speculate, for example, how in "Paul's Case," the short story by Willa Cather, the author makes use of symbols.

Level 4 questions, like the one below, compel the student to relate what he has read to real-life situations:

One of the greatneses of Thoreau is that he anticipated by many years some of the evils that would rise in American life. What were these evils? Did he propose any remedy? If so, what?²

Finally, we come to Level 5, which differs radically from Level 4 in requiring evidence of a change of *behaviour* on the part of the student. Consequently, Level 5 is at once the most important level and the most difficult one to evaluate, as Gordon indicates:

The last area of "testing" is the place to determine whether education has or has not taken place. Where there is no real change in attitude, "education" may well be dealing in mere verbalisms. It matters very little how much the pupil can say the "right" things in class, or on attitude tests; the real test is his behaviour. Where does he stand when a group is baiting an unpopular student? He may talk fluently about the dignity of the individual, but shout down others at a class meeting. He may intellectually recognize the fallacy of Emma Bovary's thinking, but live in his own world of unreality.³

To return to Level 1, it is, in Gordon's opinion, the lowest, or least "intellectual," form of testing because it merely calls for the student to reproduce something he has read or heard. Yet, according to some authorities, this form of testing is all too prevalent in English. As Clarence W. Hach asserts in the *Illinois English Bulletin* of February, 1965:

Too often we still give final tests that are almost entirely factual. If they go beyond the factual at all, they probably include a generalization that the teacher or someone else has made about the piece of literature and ask that the pupils discuss it. Such a question on *The Adventures of Huckleberry Finn* might be: "During the course of the story *The Adventures of Huckleberry Finn*, Huck changed from a carefree boy to a thoughtful young adult. Mention three experiences he had and tell how each helped him gain a better understanding of life."⁴

Like Gordon, he points to the student's change of behaviour as the crucial evidence of successful teaching:

If the emphasis in modern education is to be on altering behavior patterns, not on memory and the acquisition of knowledge *per se*, then our major tests in literature must

be of the type that requires pupils to use knowledge and demonstrate their control over it. Our major tests in literature must help us to assess pupils' growth in reading; they must help them to grow in interpreting literature and in applying knowledge of literature to life and their own behavior, and enjoy literature, to prefer a really good poem to a poor one, to want to read, most of the time at least, novels and short stories that deal with the truth of human experiences. The phrase "altering behavior patterns" as a goal of modern education is not meant in the teaching of literature to be entirely utilitarian or moralistic or didactic, of course, for literature is important, too, on a purely esthetic level.⁵

As evidence of "bad" testing closer to home, we have the study directed last year by Professor Geoffrey Mason of the University of Victoria. He classified, according to six levels of testing, the questions appearing on the June 1966 high school examinations set by the British Columbia Department of Education. His levels, entitled Knowledge, Comprehension, Analysis, Application, Synthesis, and Evaluation, correspond roughly to Gordon's levels, except that Mason makes no reference to a change of behavior on the student's part.⁶ In classifying the questions thus, Mason's stated purpose was

to determine the extent to which simple knowledge and comprehension at the recall level, rather than objectives at the higher levels of cognition, were being measured.⁷

Here, in part, is his appraisal of the data he collected:

On the English 40 examination 50% of the literature marks was allocated to the compulsory questions. This 50% was divided between Levels 1, 2 and 3 — 49.1%, and Level 4 — 0.9%. The remaining 50% of the marks was given to questions for which alternatives were provided. For example, four questions representing 8.9% of the total marks had to be selected from six questions, four of which were at Level 3, and one each at Levels 4 and 6. *An inspection of the data reveals that no question above Level 3 need to have been selected in writing the examination . . .* The Table shows that English 91 relied heavily on the lowest level, the recall of specific facts. No question beyond Level 3 was found for this examination. English 100 utilized Levels 1, 2, and 3 extensively. There were no questions at Levels 4 or 5.⁸

Mason's sobering summary of his findings reads as follows:

The data from this study support to some degree the criticism that the Departmental examinations emphasize the

recall of specific information. While the level of analysis has also received considerable attention on the 1966 English examinations, the higher levels . . . have been almost uniformly ignored.⁹

Although conventional written tests are obviously inadequate tools for measuring behavioural change (Level 5) in a student, such tests can be employed to assess abilities at the other four levels. A fundamental requirement is that the teacher must first settle upon the specific objectives to be tested and then decide what kinds of questions he will pose to test those objectives. In his excellent pamphlet on testing in English, Carruthers suggests this procedure:

The teacher must first (1) review his objectives for the unit. The test should measure *all* of the important objectives of the unit. For example, for our sample unit for English 10, "Understanding Human Nature," the teacher may have for one of his objectives "To enable the students to become more aware of characters' motives." The teacher should review this objective and the steps he took to attain it. He should then (2) decide whether this objective can be effectively tested in a paper and pencil test, and (3) decide what kind of written evidence will reflect the student's having attained the objective. He decides that it can be tested and concludes that such specific evidence as this will suffice: "Awareness of the effect of a character's actions upon others"; "awareness of the implications of a character's conversations with others"; and "understanding of the inner conflict of a character." Thus, he has sharpened his *general* teaching objective ("To enable the students to become more aware of characters' motives") to the specific *testing* ones ("Awareness of the effect of a character's actions upon others . . .", etc.) He must then in the same manner point up and sharpen each of the other general objectives for the unit.¹⁰

The next step for the teacher is to determine whether the "weight" which he gave to certain objectives when teaching the unit is fairly reflected in the test:

In the interests of fairness and validity, he will make sure, for example, that if in our sample unit he spent about 50% of the time — and emphasis — of the unit on the study of literature and its outcomes, he will allocate on the test a like weight (50 per cent) for items in the area of literature and its outcomes.¹¹

He needs also to determine whether his questions are thoughtfully distributed over the four levels of intellectual activity referred to

above. Not until this point is the teacher ready to turn his attention to the actual questions, or test items, he will employ. The kinds of decisions he will face are illustrated as follows:

[In order] to "test the student's ability to spell words encountered in the unit," an item for correction, in which the student *corrects* a misspelled word, is probably more valid for the purpose than a four-choice item in which the student is required merely to *select* a misspelled word but not to spell it correctly. (He may know which one of the four choices is misspelled, but he may not be able to *spell* it correctly.) Or, to test the student's "ability to use a comma before (certain) coordinating conjunctions which join two independent clauses," a short exercise which reads "Punctuate correctly each of the following sentences" is probably more valid for the purpose than an item which reads, "Make a list of the uses of a comma."¹²

The teacher must also decide when to employ essay items or short-answer items, which include completion, true-false, and matching items, and a wide range of multiple choice items. In a lengthy and helpful discussion of the uses and limitations of short-answer items, Carruthers warns against completion items for testing in English, because they test chiefly the student's ability to recall facts (Level 1). True-false items are ruled out, except, possibly, when the teacher wishes "to test misconceptions which the student may have in a given area of English, particularly before the beginning of the study of that area." "Matching" items are criticized for their "limited flexibility." Multiple choice items, on the other hand, can be applied to a wide variety of testing situations, ranging from simple recall of facts to more complex intellectual operations like forming judgments, or responding to the author's techniques.¹³ A few literature items will serve as illustrations:

- (1) In *The Thread That Runs So True*, the author describes his experiences as a (1) young schoolteacher (2) minister's son (3) tailor's apprentice (4) track champion.
- (2) In *Julius Ceasar*, Brutus's inner conflict is shown in the line (1) "Speak, hands, for me!" (2) "Cowards die a thousand deaths . . ." (3) "Not that I loved Caesar less . . ." (4) "he doth bestride the narrow world like a Colossus."
- (3) In *Dear Brutus*, the experiences of the persons in the play suggest that (1) some men are influenced more by evil than by good (2) happiness can be too dearly bought (3) human nature is ever the same (4) patriotism is an ever present need of man.
- (4) The plays of Sir James M. Barrie are characterized by (1)

blank verse (2) "choruses" to accentuate action (3) unhappy ton took refuge in (1) drinking (2) eating excessively endings (4) humorous stage directions.¹⁴

That pitfalls abound for the unwary in the construction of such items is illustrated by the "poor" and "improved" examples quoted below:

Poor: His early career having been atrophied, Sydney Carton (3) stealing (4) hoarding money (5) fleeing to the country.

(This item presents two vocabulary problems: "atrophied" and "took refuge in." Such expressions may be too difficult for the student.)

Improved: A habit of Sydney Carton's was (1) drinking (2) eating excessively (3) stealing (4) hoarding money (5) fleeing to the country.

Poor: Robert Louis Stevenson was very well known as an author of (1) plays (2) novels (3) chronicles of King Arthur (4) literary criticism (5) poetry.

(Although the teacher intended that number 2 be the correct answer, number 5 is also correct.)¹⁵

Poor: (1) Stowe (2) Cabell (3) Cooper (4) Poe (5) Norris wrote *Uncle Tom's Cabin*.

(In the above item, a student must read through five alternatives before he knows what the item is about or exactly what the teacher wishes to know.)

Improved: Who wrote *Uncle Tom's Cabin*? (1) Stowe (2) Cabel (3) Cooper (4) Poe (5) Norris.¹⁶

Turning to essay items, Carruthers praises their appropriateness in situations where the teacher of English

wants to measure the student's ability to organize *his own* ideas, his ability to write, to analyze, to judge, to discriminate, to give his own reaction to a work of literature, to create his own literary work, to tell his own experience, all in *his own words*.¹⁷

The construction of essay items brings the teacher face to face with many of the same problems he confronts with short-answer items. For example, this essay item may pose a vocabulary problem for the student, if he does not know the meaning of "succinctly": "Show how the author succinctly gives his views on the subject."¹⁸ Or, an item may be poorly worded, failing to define the problem or to indicate its limits for the student, as in the following case:

Some of the selections we have read have given us an idea of America. Select two books and in each case show

what idea is presented about America.

(This question is subject to several interpretations because it is so general. It also probably lacks reliability; that is, individual students will interpret it in many ways. What phases of American life or ideals should the student discuss? To what extent?)

Here is an improved version of the question:

In some of the selections we have read during this past year, the authors have informed us about America: its ideals, its customs or traditions, its heroes, or its contributions to mankind. Basing your discussion upon one full length biography and one book of non-fiction, which you identify by author and title, show that you have been informed about one or more of the areas mentioned.

The "Improved" version is much more specific than the "Example." A vague generalization is narrowed to aspects with which the student can deal in the time allotted.¹⁹

Examples of good questions on literature can be found in sample tests prepared by the College Entrance Examination Board (CEEB) for college-bound students and in sample Co-operative English Tests published by Educational Testing Service. The CEEB tests for grades nine through twelve are extremely useful to the teacher for two reasons. First, they indicate to him what college-bound students can be expected to achieve in each grade when asked to respond to complex literary works — poems, plays, novels, and essays. Second, the tests make it possible for the teacher to compare the standing of his students with that of other high school students. To give the greatest possible assistance to the teacher, the CEEB tests provide full-length sample answers submitted by students of varying abilities, the grades (from 1 to 5) assigned to those answers, and the examiners' explanations of how they arrived at the grades.

The following sample items from the CEEB grade eleven literature tests display the desirable characteristics of good essay-type questions: they are sharply limited in scope, and they furnish the student with directions that are clear and precise, yet brief. The first sample item, consisting of short essay-type questions, is taken from the section that requires the student to write an analysis of an unfamiliar literary passage, in this case a poem entitled "Late Rising":

1. What clues to the subject matter of the poem do you find in the first six lines?
2. What purpose do the following specific details serve?
 - (a) "at six o'clock in the morning" (line 9)
 - (b) "smart" (line 10)
 - (c) "a head the color of dust" (line 11)
3. In the light of the later development of the poem, what added significance do lines 14 and 15 have?
4. What relationship is there between the tempo and sound of words and their meaning in lines 43-54. What effect is the poet trying to create?
5. Why does the tempo change in line 55?
6. Why is it important that the reader be told how the twenty-five cents (two bits) was spent?
7. Why does the author repeat the first six lines of the poem at the end?²⁰

The second sample item, consisting of a long essay question, is taken from the section that requires the student to write an essay about a literary work he has studied:

A theme which writers have examined for centuries is the achievement of self-knowledge (a person's recognition of his own strengths and weaknesses, values and prejudices, aspirations and fears). Inevitably this self-knowledge is achieved only after the person has undergone an ordeal which has forced him to re-examine his own values and those of the world about him.

Directions:

Write an essay in which you show how an author has dealt with this theme in a major work (novel, drama, long narrative poem, or biography). Identify the work and the character you are writing about and be *specific* and *thorough* in your examination of the way the writer has developed this theme. You may need to refer to the action and setting, or to symbolic elements of the work, but you should concentrate on the changes in attitude which the character undergoes. *Show that you understand the nature of his new self-knowledge.*²¹

On the other hand, the Co-operative English tests in literature consist solely of multiple choice items. As mentioned earlier, Caruthers believes that the adaptability of these items makes them the most useful of all short-answer items to the teacher of English. This adaptability is suggested in sample Co-operative English tests measuring reading comprehension and consisting in part of passages accompanied by multiple choice items; for the items range

from those requiring the student to recall a fact to those requiring him to derive interpretations from what he reads. Reproduced below are a few of the items that follow upon a narrative beginning "The day of the Pony Derby was cloudless and still, a perfect day":

- (1) The weather on the day of the Derby was A. windy B. wet
C. cold D. fair
- (2) Mrs. Krug lined up beside Bozo to
A. urge him on.
B. get some exercise.
C. keep him on the track.
D. keep Maybelle from falling off.
- (3) One way this race differed from most horse races is that
A. the slowest pony was to be the winner.
B. all the entries came close to winning.
C. a foot runner accompanied one of the entries.
D. all the ponies went around in a walk.
- (4) The business manager's note probably said that
A. there was not enough money for two \$50 prizes.
B. prizes should be given to all eleven entries.
C. Sunny should get \$100 and Bozo \$50.
D. the race should be cancelled.²³

Incidentally, the CEEB tests are used by a few Canadian universities, and a move is underway at the present time to devise similar nationwide tests in Canada. The project is being financed by a group of Canadian colleges and universities and by the provincial departments of education. With more sophisticated knowledge about testing to guide them, and with material like the CEEB samples before them, the persons responsible for the English tests will no doubt be constantly seeking ways in which to improve the rationale. A better rationale will, of course, go far beyond the mere recall of facts to test, in the fullest sense, the student's responses to literature.

REFERENCES

1. Edward J. Gordon, "Levels of Teaching and Testing," *English Journal*, XLIV (September 1955), p. 331.
2. *Ibid.*, pp. 331-33.
3. *Ibid.*, p. 334.
4. "Improving Testing in English," *Illinois English Bulletin*, Urbana Ill.: Illinois Association of Teachers of English, p. 9.
5. *Ibid.*, p. 10.

6. "A Taxonomic Analysis of the English 40, English 91 and English 100 Departmental Examinations," *B.C. English Teacher*, 7 (March 1967), pp. 21-22.
7. *Ibid.*, p. 21.
8. *Ibid.*, p. 23. Italics added.
9. *Ibid.* It may or may not comfort teachers of English to know that much the same situation has been reported concerning testing in history. The tests set by the State Board of Regents of New York came under fire last Spring because they emphasize the recall of facts. After commenting that the tests "[foster] the type of educational emphases generally conceived to be most deadly to classroom interest and student improvement," the writer quotes Gilbert Highet's observation to the effect that the student is forced to see the subject "as a congeries of unrelated little facts . . ." (Henry M. Littlefield, "Who's Afraid of Regents Exams?" in *Teachers College Record*, 68, March 1967, 481).
10. *Building Better English Tests: A Guide for Teachers of English in the Secondary Schools*, Champaign, Ill.: National Council of Teachers of English, 1963, p. 6.
11. *Ibid.*, p. 7.
12. *Ibid.*, p. 9.
13. *Ibid.*, pp. 14-18.
14. *Ibid.*, p. 21.
15. *Ibid.*, p. 22.
16. *Ibid.*, p. 24.
17. *Ibid.*, p. 25.
18. *Ibid.*, p. 27.
19. *Ibid.*, pp. 26-27.
20. "End-of-Year Examinations in English for College-Bound Students, Grades 9-12," Princeton, N.J.: College Entrance Examination Board, 1963, pp. 7-8.
21. *Ibid.*, p. 9.
22. "Form 2A: Reading Comprehension," *1960 Co-operative English Tests Specimen Set*, Princeton, N.J.: Educational Testing Service, pp. 4-5.